AWIPS-2 Application Focal Point Course

Archiver Configuration Exercise for Local and Manual Data

Warning Decision Training Branch

National Weather Service Training Division

AWIPS-2 Build OB14.2.4

December 2014

Table of Contents

Archiver Configuration Exercise for Local and Manual Data

Objective. Verify that products ingested through LDAD and/or the manual endpoint are correctly archivable by appearing in the archiver GUIs, both for Processed and Raw data. If products do not appear in the GUIs, make required changes to the archiver configurations.

Introduction. Accessible through the Localization Perspective, the archiver configuration files (RAW_DATA.xml and PROCESSED_DATA.xml) define datasets and specify how to obtain date and time information for each archivable file. These archivable files are located in /data_store for RAW_DATA.xml and /archive for PROCESSED_DATA.xml; each of these XML files contains an entry that defines its root directory for the files it handles. The date and time information from the dataset definitions is used by the archiver to determine which data to save during case creation and which data to purge based on retention times. The process we need to follow here to configure the datasets is to
(1) analyze the structure of how the data files are stored (directories and filenames),
(2) analyze the corresponding entries in the archiver configuration file (RAW_DATA or PROCESSED_DATA) to see if they collect the archivable files, if they need to be modified, or if brand-new entries are required;
(3) make any corrections to the configuration and verify the configuration takes effect using the Case Creation or Retention GUIs.

This exercise shows multiple examples of this process for both raw data in /data_store/manual and processed data in /archive. An Appendix is provided for reference at the end of the exercise for more file and directory structures observed at additional AWIPS sites. These structures sometimes use additional ways to specify the date and time, including the Julian date.

Note: This exercise was developed using the baseline RAW_DATA.xml and PROCESSED_DATA.xml files from Build OB14.2.4. These versions of these particular baseline RAW_DATA.xml and PROCESSED_DATA.xml files handle some situations that had to be included in site overrides prior to this build. In particular, if your office has completed this exercise prior to Build OB14.2.4, you should remove your local SITE's override entries in RAW_DATA.xml that handled MPE, HPE, LAPS/MSAS, SHEF, and local text and warning products.

Process Overview

Step 1. Analyze the structure of how data files are stored (directories and filenames). To do this, check the storage locations of files (/data_store/manual or /archive) and examine patterns of the filename structure for various types of data. You will need to determine distinguishing characteristics of the data and the various date/time specifications of the data. Different types of data in the manual endpoint are often comingled together, and any distinguishing features in the filenames allow the archiver to separate the data (for example, Canadian model data vs. the HRRR vs. LAPS and MSAS).

Often, especially for raw data in the manual endpoint, filenames contain multiple date/time specifications. These times include when the manual endpoint received the data, a valid time, the issuance time of the data/product, the time of the model run cycle, or the time when the LDAD system received the data. EDEX, when it finds data in the manual endpoint, organizes data in /data_store/manual by subdirectories according to a decoder plugin that can process it and then by date and hour directories of when the data were received by the manual endpoint for processing. For example, a nine-hour forecast grid of surface pressure from a 12Z run of the Canadian model was stored as /data_store/manual/grib/20140304/15/LDAD-GRIB.CMC2_reg_PRES_SFC_0_ps15km_2014030412_P009.Grb2_2014-03-04_150055.

There are trade-offs involved in determining which time to use. From an accuracy perspective, it is better, if possible, to use a valid time from the filename itself rather than from the directory structure. In the example above, the directory is the date/time when the grid was received (15Z), not the valid time or reference time of the model (12Z). In some cases, the receipt time is all we have to work with. However, the performance of the archiver GUIs is significantly improved when the software can use the date/time from the directory rather than have to parse it out of every filename, of which there could be tens or hundreds of thousands or more on the system at once. Thus, from the performance perspective, the directory date/time is preferred.

Step 2. Analyze the appropriate sections of RAW_DATA.xml and PROCESSED_DATA.xml to verify if the existing entries match the structures actually present in the corresponding directories. These entries contain patterns that use regular expressions, and can be difficult for humans to read. To help simplify this for you, here are some common expressions that are used often in RAW_DATA.xml and PROCESSED_DATA.xml with their meanings:

Pattern or symbol	Meaning
.	Picks up and matches any one character (a one-character wildcard).
\.	Matches a dot character (.). The backslash "escapes" the dot character, so in this case it is a literal dot rather than a one-character wildcard.
*	The previous expression(s) any number of times. So .* matches any characters in the filenames, depending upon what precedes or follows the .* in the pattern.
\d{n}	Matches n digits, 0 through 9. \d{4} matches four digits (usually for YYYY) and \d{2} picks up two digits. \d{10} matches ten digits.
()	The parentheses are not part of normal regular expression nomenclature. AWIPS-2 uses parentheses here to define a regular expression "group". The groups are numbered sequentially and used by other parts of the dataset definition (namely, the displayLabel and dateGroupIndices). This is a similar methodology to the groupings used in LDM's pqact.conf.
Regular character	Matches itself. This includes the forward slash (/) for directory separation, the dash (-), and underscore (_) which are used in various filenames, plus other regular alphanumeric characters.
[]	Used in regular expressions for internal grouping (not the same as the regular parenthesis described above)

^	An inverse operator. [^\d] means “any non-digit character”.
|	The pipe is an alternative operator. z|\d{2}z means find either a z character or two digits followed by a z character.

Another good online reference for regular expressions is available from oracle.com:
http://docs.oracle.com/javase/tutorial/essential/regex/index.html

Here are the entries in the base version of RAW_DATA.xml for /data_store/manual, as of OB14.2.4. In the explanatory boxes below, the # sign represents a digit from 0-9, and in many cases ten # signs represents a time in Epoch seconds (the number of seconds that has elapsed since 1/1/1970). In some other particular cases, ten # signs in a filename represent a time in YYYYMMDDHH format. The asterisks in the boxes below represent a standard filename wildcard.

❶

```
<dataSet>
    <dirPattern>(manual)/grib/(\d{4})(\d{2})(\d{2})/(\d{2})/</dirPattern>
    <filePattern>.*(LAPS|MSAS).*</filePattern>
    <timeType>Date</timeType>
    <displayLabel>{1}-LAPS/MSAS</displayLabel>
    <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

matches *LAPS* and *MSAS*
/data_store/manual/grib/20140421/15/
ZETA98.LAPS.20140421_1500

The above entry defines YYYY, MM, and DD (groups 2, 3, and 4) to come from the directory names in /data_store/manual/grib. HH (group 5) is an hourly subdirectory of the YYYYMMDD directory. Inside the hourly directory, this entry finds all files that include “LAPS” or “MSAS” as part of their filename.

❷

```
<dataSet>
    <dirPattern>(manual)/grib/(\d{4})(\d{2})(\d{2})/(\d{2})/(mpe)</dirPattern>
    <filePattern>.*_(\d{4})(\d{2})(\d{2})(\d{2})z.*</filePattern>
    <timeType>Date</timeType>
    <displayLabel>{1}-MPE Local</displayLabel>
    <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

matches *_##########z.*
/data_store/manual/grib/20140414/21/mpe/
ZETA98_2014041421z_14212507.grib

The above entry was designed for files in the mpe subdirectory of manual/grib/YYYYMMDD/HH/. It picks up files that have an underscore just prior to a set of 10 digits followed by a “z”. While the ten digits do represent a date/time, the actual date/time values defined by this entry come from the <dirPattern> (Groups 2, 3, 4, and 5) and not the <filePattern> (which would have been Groups 7, 8, 9, and 10).

❸

```
<dataSet>
    <dirPattern>(manual)/grib/(\d{4})(\d{2})(\d{2})/(\d{2})/(mpe)</dirPattern>
    <filePattern>(ZETA98_)[A-Z].*(\d{4})(\d{2})(\d{2})(\d{2})(\d{2})z.*</filePattern>
    <displayLabel>{1}-HPE/Bias HPE</displayLabel>
    <timeType>Date</timeType>
    <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

matches ZETA98_X*############z*
(X represents any number of capital letters)
/data_store/manual/grib/20140414/21/mpe/
ZETA98_DHRMOSAIC201404142158z_21170324.grib

The entry above likewise was designed for files in the mpe subdirectory of manual/grib/YYYYMMDD/HH. It picks up files starting with "ZETA98_" (just like the previous entry), but this one must have additional capital letters (A-Z) prior to 12 digits followed by a z.

❹

```
<dataSet>
   <dirPattern>(manual)/.*taf/(\d{4])(\d{2})(\d{2})/(\d{2})</dirPattern>
   <displayLabel>{1}-Local TAFs</displayLabel>
   <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

matches all files in /data_store/manual/*taf/YYYYMMDD/HH

The previous entry matches any files in date/hour subdirectories of any taf directory under /data_store/manual. Sometimes locally-produced TAFs are stored in either /data_store/manual/taf or /data_store/manual/nctaf, so this entry allows for both.

❺

```
<dataSet>
   <dirPattern>(manual)/shef/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <displayLabel>{1}-Various SHEF Products</displayLabel>
   <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

matches all files in /data_store/manual/shef/YYYYMMDD/HH

This entry is similar to the previous TAF entry. It matches any file stored in a YYYYMMDD/HH subdirectory in /data_store/manual/shef.

❻

```
<dataSet>
    <dirPattern>(manual)/text/\d{8}/\d{2}</dirPattern>
    <filePattern>.*(\d{10})</filePattern>
    <displayLabel>{1}-Various Local Text Products</displayLabel>
    <timeType>EpochSec</timeType>
    <dateGroupIndices>2</dateGroupIndices>
</dataSet>
```

matches *.########## in /data_store/manual/text/########/##

❼

```
<dataSet>
    <dirPattern>(manual)/warning/\d{8}/\d{2}</dirPattern>
    <filePattern>.*(\d{10})</filePattern>
    <displayLabel>{1}-Local Warnings</displayLabel>
    <timeType>EpochSec</timeType>
    <dateGroupIndices>2</dateGroupIndices>
</dataSet>
```

matches *.########## in /data_store/manual/warning/########/##

These last two entries (❻ and ❼) were designed for files that are located in /data_store/manual/text or /data_store/manual/warning, respectively, and have their valid time expressed in Epoch seconds (the number of seconds that have elapsed since midnight on January 1, 1970) as a ten-digit number ending the filename.

In summary, these baseline entries in RAW_DATA.xml were designed to handle LAPS and MSAS analyses from /data_store/manual/grib, HPE and MPE files from /data_store/manual/grib/mpe, local SHEF

products, and locally-generated TAFs, warnings, and text products. It is possible (perhaps, even likely) that your local system may place additional files in the directories defined in these entries (/data_store/manual/grib/, /data_store/manual/grib/*/*/mpe, /data_store/manual/text, and /data_store/manual/warnings) that are not represented by the specific <filePattern> tags. Thus, you should still examine these directories to verify all files there will be picked up by the archiver.

Step 3. Make any required corrections to the configuration and verify the configuration takes effect using the Case Creation or Retention GUIs. You will need to construct entries in RAW_DATA.xml or PROCESSED_DATA.xml that will conform to your actual file listings. In some cases, you can modify existing dataset definitions. **You will need to use the Localization Perspective to make a site override of RAW_DATA.xml or PROCESSED_DATA.xml to include your new entries.** In the Localization Perspective, these archiver configuration files are located under Archive ► Configuration. **If you have not yet made a site override for these files, you will need to add permission to your user for the Localization Perspective to create and edit these files.** The permission is com.raytheon.localization.site/common_static/archiver/purger. You should use the AWIPS User Administration GUI to add this permission and to assign this permission to specific users. Step-by-step instructions on using the AWIPS User Admin GUI to add this permission are provided in Appendix 1.

The BASE versions of the archiver configuration files are relatively well-documented internally. Each dataset definition requires a <dirPattern> tag and a <displayLabel> tag with an optional <filePattern> tag along with other specifications. The <dirPattern> and <displayLabel> tags determine the appearance of each dataset in the Archiver GUIs. The <dirPattern> tag, along with the <dateGroupIndices> tag, also provides a mapping to the date/time information in the actual filename and directory structure. In some cases, there may be no appropriate date/time information in the directory but only in the filenames themselves, so the <filePattern> tag additionally must be used.

The <displayLabel> tags which generate the data listings in the Retention and the Case Creation GUIs only work with the regular expression groups that are defined in <dirPattern>, not with those in <filePattern>. Thus, when file patterns are necessary to discriminate between files in a single directory, this exercise creates explicit labels, so the GUI tells you exactly what you are archiving. The downside to this approach is that other potentially archivable files that do not match the pattern will not get archived unless you used a more generic file pattern. For example, a very generic pattern like `<dirPattern>(manual)/.*/\d{8}/\d{2}</dirPattern>`, would archive most data in the manual endpoint but with little ability to choose which datasets to save or to not save.

Important note on /data_store/manual organization. Lots of different kinds of data are ingested into AWIPS-2 using the manual endpoint. The system attempts to organize these data automatically by plugin. Nevertheless, this rough organization still allows different types of data to be located in the same directories. For example, any particular manual/grib/<YYYYMMDD>/<HH> directory may contain output from multiple models and analyses (e.g., Canadian, HRRR, LAPS, MSAS, etc.). The archiver configuration allows you to specify and archive each of these independently, which is the approach followed in this document. In order to do this, however, you need to identify some distinguishing

characteristic of the filename patterns for a given model. Where appropriate, these are identified in each section below.

Modifying PROCESSED_DATA.xml is usually only required to handle those situations where local data is processed by a new or "non-baseline" plugin. This is not a common occurrence because most local data processed by the manual endpoint is actually ingested by the grid, text, or some other standard plugin, and thus is stored in the file structure that plugin normally uses. Among the examples given below in this exercise, the only data from the manual endpoint that is processed by its own plugin is sportlma. Therefore, there is only one example of configuring PROCESSED_DATA.xml. An additional example of needing to configure PROCESSED_DATA.xml is presented in Appendix 2; it is for the regionalsat plugin.

Application to Actual Archivable Files

This section of the exercise applies the three steps outlined above to 12 specific and common examples.

1. Raw Data: SPoRT LMA (Lightning Mapping Array)
2. Processed Data: SPoRT LMA (Lightning Mapping Array)
3. Raw Data: Half-Degree GFS (GFS230) Model
4. Raw Data: LAPS and MSAS
5. Raw Data: Local MPE (Multi-Sensor Precipitation Estimator)
6. Raw Data: HPE (High-Resolution Precipitation Estimator)
7. Raw Data: Canadian Model
8. Raw Data: HPC Winter Weather Desk Products
9. Raw Data: TAFs
10. Raw Data: SHEF (Standard Hydrometeorological Exchange Format) Products
11. Raw Data: Local Text Products
12. Raw Data: Local Warnings

Additional examples are included in Appendix 2. You can pick which of these examples to follow. They are applicable to a variety of situations across the NWS.

Data Example 1. Raw Data: SPoRT LMA (Lightning Mapping Array)

Step 1. Check /data_store/manual and examine the filename patterns for the sportlma data stored there. Analyze the filenames to determine the format of date/time stamps.

```
manual/sportlma/20131118/16/sportlma_hglma_fid_20131118_1608
manual/sportlma/20131118/16/sportlma_pgok_fed_20131118_1607
manual/sportlma/20131118/16/sportlma_colma_fed_20131118_1608
manual/sportlma/20131118/16/sportlma_pghg_fed_20131118_1608
manual/sportlma/20131118/16/sportlma_pgok_fid_20131118_1608
manual/sportlma/20131118/16/sportlma_pgdc_fid_20131118_1606
manual/sportlma/20131118/16/sportlma_lllma_fed_20131118_1606
manual/sportlma/20131118/16/sportlma_pghg_fid_20131118_1608
manual/sportlma/20131118/16/sportlma_nalma_rfd_20131118_1606
```

```
manual/sportlma/20131118/16/sportlma_colma_rfd_20131118_1608
manual/sportlma/20131118/16/sportlma_pghg_fid_20131118_1607
manual/sportlma/20131118/16/sportlma_pgna_fed_20131118_1606
```

The date/time stamps for the valid times of the SPoRT LMA data are obviously at the end of each filename, but the directory structure also contains the date/time for the receipt of the data. In most cases, these two times agree for diagnostic data like these. The filename structure can be described as: **manual/sportlma/YYYYMMDD/HH/sportlma_{product-dependent-information}_YYYYMMDD_HHMM.**

In the file listing above, you will notice the product dependent information specifies the domain/sensor and product. For the domain and sensor, "nalma" means "Northern Alabama Lightning Mapping Array" and "pgok" means "pseudo-GLM [Geostationary Lightning Mapper] for Oklahoma". For the type of product, "fed", "fid", "mfd", "rfd", and "sd" are various types of flash densities and "fi" is "flash initiation". Other LMA domains include:

- co: Colorado
- dc: District of Columbia
- hg: Houston
- ll: Florida
- wt: West Texas

Step 2. Analyze existing patterns in RAW_DATA.xml. SPoRT LMA data is new data and is not represented in the baseline RAW_DATA.xml file.

Step 3. Make any required corrections to the configuration and verify the configuration takes effect using the Case Creation or Retention GUIs.

This dataSet entry includes the sportlma directory and uses the date and hour directories in the <dirPattern>. A second version is also provided to illustrate how to extract certain domains or sensors from this data. Because the regular expression syntax is challenging to read, the matches between the pieces of the regular expression, the groups, and corresponding directories and filenames are highlighted in color. The colors below are given in the same order but do not always represent the same regular expression syntax. This entry was constructed using the performance approach, matching the date/time from the directories.

```
<dataSet>
   <dirPattern>(manual)/(sportlma)/(\d{4})(\d{2})(\d{2})/\d{2}</dirPattern>
   <timeType>Date</timeType>
   <displayLabel>{1}-{2}</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

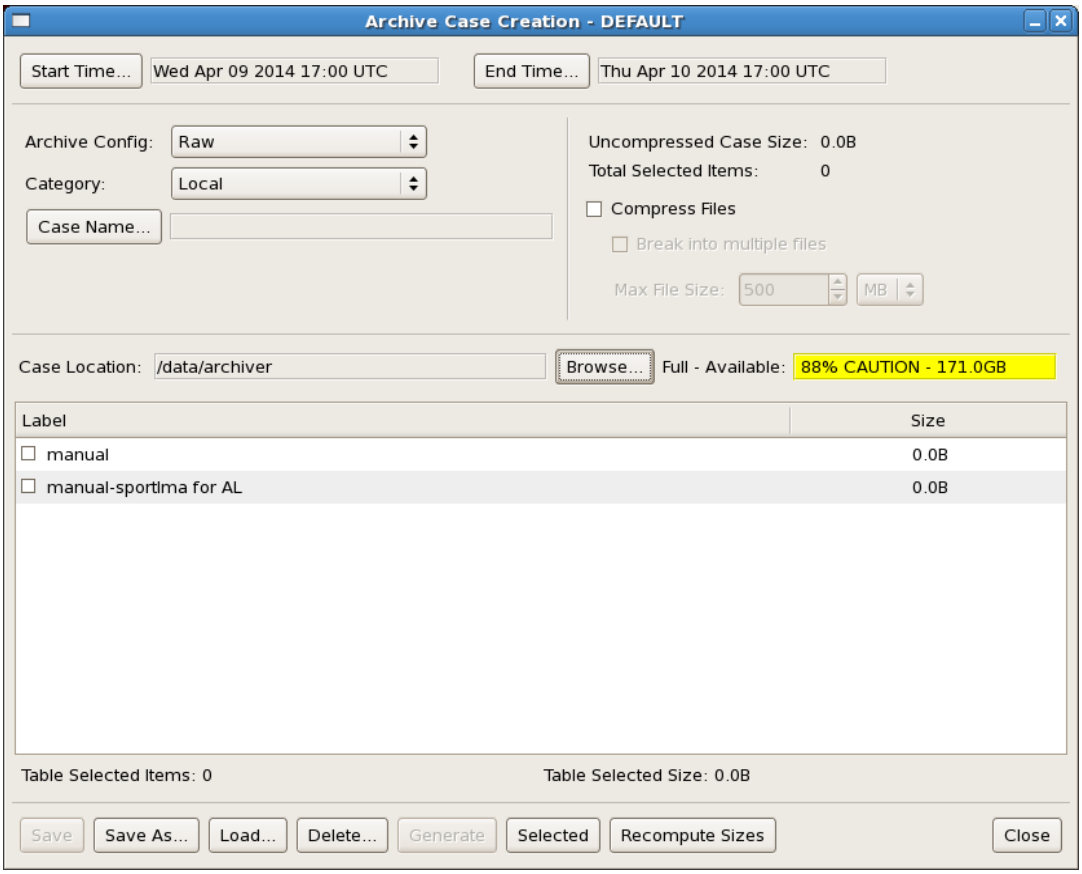
Sample file: manual/sportlma/20131118/16/sportlma_nalma_fid_20131118_1608

Second version for domain specification (in this case, for the Alabama domain):

```
<dataSet>
   <dirPattern>(manual)/(sportlma)/(\d{4})(\d{2})(\d{2})/\d{2}</dirPattern>
   <timeType>Date</timeType>
   <filePattern>.*(nalma|pgna).*</filePattern>
   <displayLabel>{1}-{2} for AL</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataset>
```

Sample file: manual/sportlma/20131118/16/sportlma_nalma_fid_20131118_1608

If you add this dataSet definition to the Local category in RAW_DATA.xml and have the appropriate data in /data_store/manual, the result should be a new line in the Case Creation GUI (or Retention GUI as well) in the Local category of Raw data that is labeled “manual-sportlma”:



Data Example 2. **Processed Data: SPoRT LMA (Lightning Mapping Array)**

Step 1. Analyze the file/directory structure. Here is a sample directory listing of a few hours of processed data from /archive/sportlma:

```
COLMA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
COLMA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
COLMA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
COLMA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
COLMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
COLMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
COLMA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
COLMA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
COLMA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
COLMA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
COLMA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
COLMA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
COLMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
COLMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
COLMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
COLMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
COLMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
COLMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
DCLMA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
DCLMA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
DCLMA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
DCLMA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
DCLMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
DCLMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
DCLMA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
DCLMA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
DCLMA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
DCLMA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
DCLMA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
DCLMA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
DCLMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
DCLMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
DCLMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
DCLMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
DCLMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
DCLMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
LLLMA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
LLLMA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
LLLMA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
LLLMA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
LLLMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
LLLMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
LLLMA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
LLLMA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
LLLMA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
LLLMA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
LLLMA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
LLLMA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
LLLMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
LLLMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
LLLMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
LLLMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
LLLMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
LLLMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
NALMA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
NALMA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
NALMA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
NALMA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
NALMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
NALMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
NALMA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
NALMA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
NALMA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
NALMA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
NALMA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
NALMA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
NALMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
NALMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
NALMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
NALMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
NALMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
NALMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGCO/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGCO/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGCO/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGCO/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGCO/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGCO/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGCO/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGCO/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGCO/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGCO/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGCO/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGCO/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGCO/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGCO/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGCO/mfd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGCO/mfd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGCO/mfd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGCO/mfd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGNA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGNA/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGNA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGNA/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGNA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGNA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGNA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGNA/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGNA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGNA/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGNA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGNA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGNA/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGNA/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGNA/mfd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGNA/mfd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGNA/mfd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGNA/mfd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGOK/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGOK/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGOK/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGOK/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGOK/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGOK/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGOK/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGOK/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGOK/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGOK/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGOK/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGOK/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGOK/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGOK/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGOK/mfd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGOK/mfd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGOK/mfd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGOK/mfd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGWT/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGWT/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
```

```
PGWT/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGWT/fed/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGWT/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGWT/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGWT/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGWT/fid/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGWT/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGWT/fid/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGWT/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGWT/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
PGWT/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
PGWT/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
PGWT/mfd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
PGWT/mfd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
PGWT/mfd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
PGWT/mfd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
WTLMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
WTLMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
WTLMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.bin.1
WTLMA/sd/sportlma-2014-04-07-19/sportlma-2014-04-07-19.h5
WTLMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
WTLMA/sd/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
```

In this listing, you will notice there is a second-order directory hierarchy. This plugin organizes its processed data such that the sensor/domain and product are specified in the directory structure and is similar to the raw data.

Step 2. Analyze existing patterns in PROCESSED_DATA.xml. This data is not represented in the baseline PROCESSED_DATA.xml file.

Step 3. Modify PROCESSED_DATA.xml to incorporate this data, and verify the modifications take effect using the Case Creation GUI.

Using the Localization Perspective, a new entry needs to be added to a site override of PROCESSED_DATA.xml to account for this directory hierarchy. In this example, we will add this data to the Local category, though it may make just as much sense to add it to the Observation category.

Before we make this change, the local category in the baseline version of PROCESSED_DATA.xml looks like this:

```
<category>
   <name>Local</name>
   <extRetentionHours>168</extRetentionHours>
   <dataSet>
       <dirPattern>(ldadhydro|ldadmesonet|ldadprofiler|ldad_manual|qc)/.*(\d{4})-
(\d{2})-(\d{2})-(\d{2}).*</dirPattern>
       <displayLabel>{1}</displayLabel>
       <dateGroupIndices>2,3,4,5</dateGroupIndices>
   </dataSet>
</category>
```

The additional dataset for this category in PROCESSED_DATA.xml for sportlma should look like this:

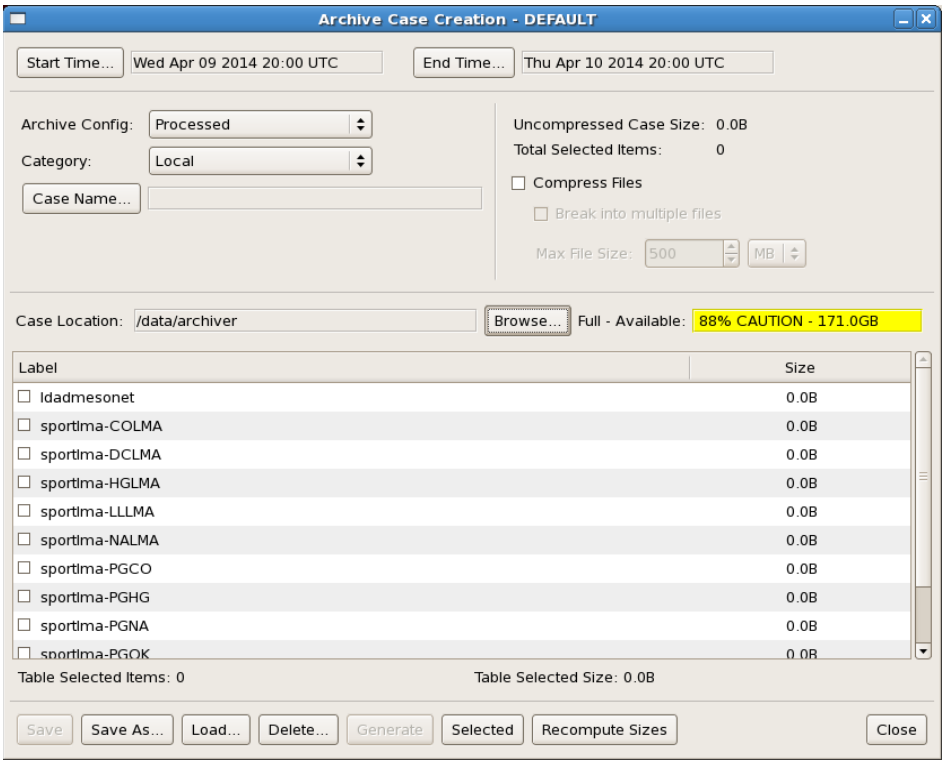
```
       <dataSet>
           <dirPattern>(sportlma)/(.*)/.*/.*-(\d{4})-(\d{2})-(\d{2})-
(\d{2})</dirPattern>
           <timeType>Date</timeType>
           <displayLabel>{1}-{2}</displayLabel>
           <dateGroupIndices>3,4,5,6</dateGroupIndices>
       </dataSet>
```

Sample Files:

```
sportlma/PGOK/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
sportlma/PGOK/fed/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
```

With no <filePattern> tag, this dataset matches every file in the /archive/sportlma/*/*/*-YYYY-MM-DD-HH directories. For the sample data given above, this dataset entry produces checkboxes in the Retention and Case Creation GUIs for each of these data types:

sportlma-COLMA
sportlma-DCLMA
sportlma-LLLMA
sportlma-NALMA
sportlma-PGCO
sportlma-PGNA
sportlma-PGOK
sportlma-PGWT
sportlma-WTLMA



If you wanted to restrict the archiving to particular sensors in certain domains, say NALMA and PGNA (Northern Alabama) for WFO HUN, the corresponding entry would be:

```
    <dataSet>
```

```
          <dirPattern>(sportlma)/(NALMA|PGNA)/.*/.*-(\d{4})-(\d{2})-(\d{2})-
(\d{2})</dirPattern>
           <timeType>Date</timeType>
           <displayLabel>{1}-{2}</displayLabel>
           <dateGroupIndices>3,4,5,6</dateGroupIndices>
      </dataSet>
```

Sample Files:

```
sportlma/NALMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
sportlma/NALMA/fed/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
sportlma/NALMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
sportlma/NALMA/sd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
sportlma/PGNA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.bin.1
sportlma/PGNA/fid/sportlma-2014-04-07-20/sportlma-2014-04-07-20.h5
sportlma/PGNA/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.bin.1
sportlma/PGNA/mfd/sportlma-2014-04-07-18/sportlma-2014-04-07-18.h5
```

Data Example 3. Raw Data: Half-Degree GFS (GFS230) Model

Step 1. Analyze the file/directory structure. Here are a few files from /data_store/manual/grib:

```
manual/grib/20131118/16/GFS3HalfDeg.13111812.gfs.t12z.pgrb2f12.grb2
manual/grib/20131118/16/GFS3HalfDeg.13111812.gfs.t12z.pgrb2f15.grb2
manual/grib/20131118/16/GFS3HalfDeg.13111812.gfs.t12z.pgrb2f18.grb2
manual/grib/20131118/16/GFS3HalfDeg.13111812.gfs.t12z.pgrb2f21.grb2
manual/grib/20131118/16/GFS3HalfDeg.13111812.gfs.t12z.pgrb2f24.grb2
```

The half-degree GFS files only have the two-digit year in the filename and thus do not contain the full date in the actual filename. The time specifications for the archive configuration XMLs do not accept a two-digit year, so the configuration will have to either use the year from the directory name, or use the file write time. In this case, the structure is **manual/grib/YYYYMMDD/HH/GFS3HalfDeg.YYMMDDHH.gfs.tHHz.pgrb2fhh.grb2**, where hh means the forecast hour. Note that the HH (that is, "16") in the directory path is the hour the file was received, and HH in the filename ("12") is the hour of the model run. From the accuracy point-of-view, the year should come from the directory and the rest of the date/time should come from the filenames.

Step 2. Analyze existing patterns in RAW_DATA.xml. The patterns in the baseline RAW_DATA.xml for /data_store/manual/grib only would handle LAPS and MSAS files (in YYYYMMDD/HH subdirectories) or particular files in /data_store/manual/grib/YYYYMMDD/HH/mpe. Because these GFS files are in /data_store/manual/grib/YYYYMMDD/HH and do not include LAPS or MSAS in the filenames, the baseline patterns don't apply.

Step 3. Modify RAW_DATA.xml to incorporate this data, and verify the modifications take effect using the Case Creation GUI.

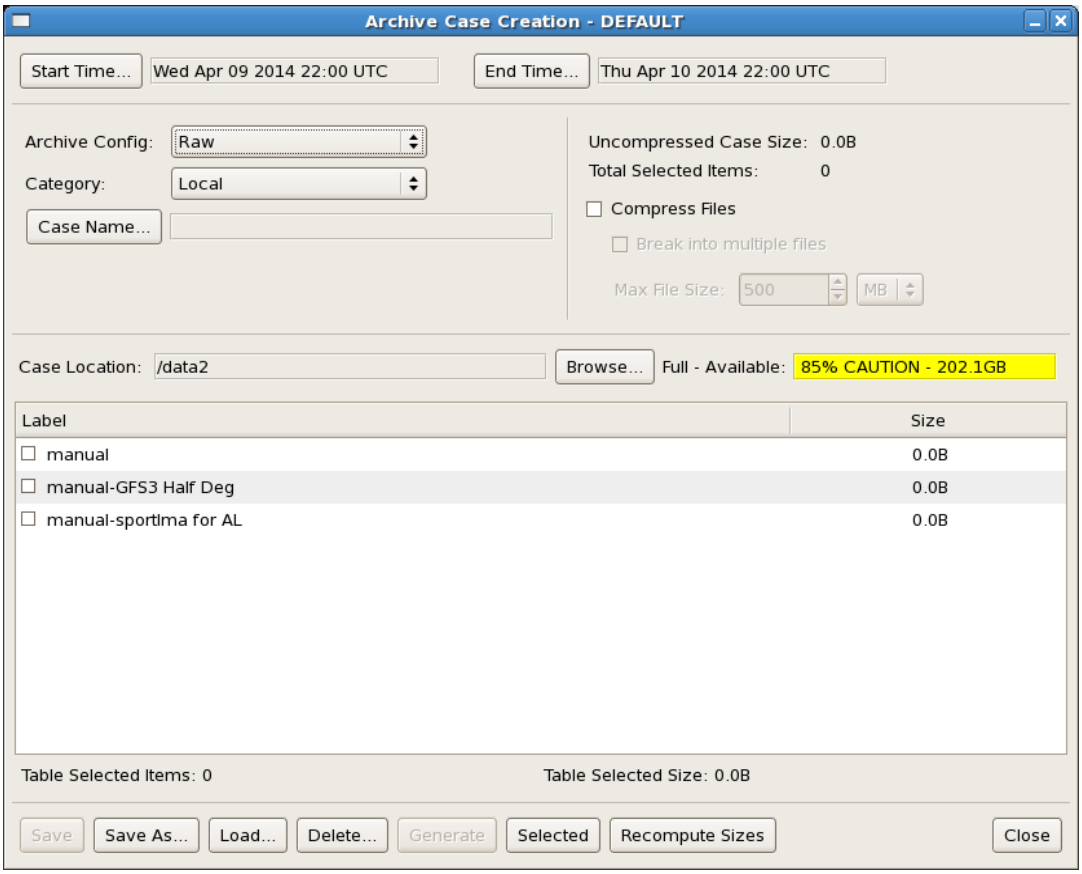
As noted above, the year needs to come from the directory and the rest of the date/time should come from the filename to ensure the correct model runs are archived. Note that any expression groups from the <filePattern> tag cannot be used in the <displayLabel>; consequently, the label explicity says “GFS3 Half Deg”. In the example below, several groups are defined for visual grouping, but are not actually used by the displayLabel or dateGroupIndices.

Distinguishing Characteristic: “GFS3HalfDeg” begins the actual filename.

```
<dataSet>
   <dirPattern>(manual)/grib/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <filePattern>(GFS3HalfDeg).(\d{2})(\d{2})(\d{2})(\d{2}).gfs.*</filePattern>
   <displayLabel>{1}-GFS3 Half Deg</displayLabel>
   <dateGroupIndices>2,8,9,10</dateGroupIndices>
 </dataSet>
```

Sample file: `manual/grib/20131118/16/GFS3HalfDeg.13111812.gfs.t12z.pgrb2f12.grb2`

The result of adding the GFS3HalfDeg entry is shown in the Case Creation GUI:



Data Example 4. Raw Data: LAPS and MSAS

Step 1. Analyze the file/directory structure. Here are a few files from /data_store/manual/grib:

```
manual/grib/20131119/02/ZETA98.LAPS.20131119_0200
manual/grib/20131119/02/ZETA98.MSAS.20131119_0200
manual/grib/20131119/03/ZETA98.MSAS.20131119_0300
manual/grib/20131119/04/ZETA98.MSAS.20131119_0400-11
manual/grib/20131119/04/ZETA98.MSAS.20131119_0400-26
manual/grib/20131119/04/ZETA98.MSAS.20131119_0400-41
manual/grib/20131119/04/ZETA98.MSAS.20131119_0400-56
```

The LAPS and MSAS data fit **manual/grib/YYYYMMDD/HH/ZETA98.MSAS.YYYYMMDD_HHMM**. MSAS (but not LAPS) is re-run multiple times per hour, but the analysis is only valid for the top of the hour. So the 20131121_1400-11 MSAS file was run at 14:11 and updates the 1400 analysis.

Step 2. Analyze existing patterns in RAW_DATA.xml. The first entry in the baseline RAW_DATA.xml for /data_store/manual/grib (❶) was designed specifically for LAPS and MSAS (in OB14.2.4), so there is no need for any further modifications for this data.

Data Example 5. Raw Data: Local MPE (Multi-sensor Precipitation Estimator)

Step 1. Analyze the file/directory structure. Below are a few MPE files from /data_store/manual/grib. Note there are also *HPE* (High-Resolution Precipitation Estimator) files in the same directory, and those are handled in the next example. Some MPE files are created locally at each WFO while others are created and disseminated by RFCs. This section deals with those created at the WFO using their local hydroapps configuration. (RFC-generated MPE files arrive at WFOs via the AWIPS SBN.)

```
manual/grib/20131119/02/mpe/ZETA98_2013111902z_19022507.grib
manual/grib/20131119/02/mpe/ZETA98_2013111902z_19032505.grib
manual/grib/20131119/02/mpe/ZETA98_2013111902z_19042503.grib
manual/grib/20131121/14/mpe/ZETA98_2013112114z_21142508.grib
```

The local MPE files follow this specification:
manual/grib/YYYYMMDD/HH/mpe/ZETA98_YYYYMMDDHHz_HHMMxxxx.grib.

Step 2. Analyze existing patterns in RAW_DATA.xml. Entry ❷ in the baseline RAW_DATA.xml (in OB14.2.4) was designed to archive MPE files. There is no need for further modification based on these filenames.

Data Example 6. Raw Data: HPE (High-resolution Precipitation Estimator)

Step 1. Analyze the file/directory structure. Here are a few HPE files from /data_store/manual/grib.

```
manual/grib/20131121/14/mpe/ZETA98_DHRMOSAIC201311211425z_21143048.grib
```

```
manual/grib/20131121/14/mpe/ZETA98_BDHRMOSAIC201311211425z_21143049.grib
manual/grib/20131121/14/mpe/ZETA98_DHRMOSAIC201311211430z_21143557.grib
manual/grib/20131121/14/mpe/ZETA98_BDHRMOSAIC201311211430z_21143559.grib
manual/grib/20131121/14/mpe/ZETA98_DHRMOSAIC201311211436z_21144108.grib
manual/grib/20131121/14/mpe/ZETA98_EBMOSAICM60201311211436z_21144110.grib
manual/grib/20131121/14/mpe/ZETA98_ERMOSAICM60201311211436z_21144116.grib
```

These files follow this pattern:
manual/grib/YYYYMMDD/HH/mpe/ZETA98_{product}YYYYMMDDHHMMz_DDHHMMxxxx.grib. The {product} for these HPE files is some type of MOSAIC. In the future, it is possible for additional files from the HPN (High-Resolution Precipitation Nowcaster) to be included in this directory (as DR17417 is addressed), and these HPN files may not be identified with the word "MOSAIC". Rather, they may be identified as PRTMXX or BPRTMXX, where XX is a forecast period in minutes.

Step 2. Analyze existing patterns in RAW_DATA.xml. An entry (❸) in the baseline RAW_DATA.xml (in OB14.2.4) was specifically designed for HPE data. There is no need for additional configurations based on the filenames listed here.

Data Example 7. <u>Raw Data: Canadian Model</u>

Step 1. Analyze the file/directory structure. Here are a few files from the Canadian model. All are located in /data_store/manual/grib:

```
manual/grib/20140304/14/LDAD-
GRIB.CMC2_reg_DEPR_ISBL_850_ps15km_2014030412_P018.Grb2_2014-03-04_145810
manual/grib/20140304/14/LDAD-
GRIB.CMC2_reg_DEPR_ISBL_700_ps15km_2014030412_P015.Grb2_2014-03-04_145815
manual/grib/20140304/14/LDAD-
GRIB.CMC2_reg_DEPR_ISBL_1000_ps15km_2014030412_P015.Grb2_2014-03-04_145815
manual/grib/20140304/14/LDAD-
GRIB.CMC2_reg_DEPR_ISBL_500_ps15km_2014030412_P012.Grb2_2014-03-04_145840
manual/grib/20140304/15/LDAD-
GRIB.CMC2_reg_DPT_TGL_2_ps15km_2014030412_P009.Grb2_2014-03-04_150035
manual/grib/20140304/15/LDAD-
GRIB.CMC2_reg_TMP_TGL_2_ps15km_2014030412_P009.Grb2_2014-03-04_150035
manual/grib/20140304/15/LDAD-
GRIB.CMC2_reg_PRES_SFC_0_ps15km_2014030412_P009.Grb2_2014-03-04_150055
manual/grib/20140304/15/LDAD-
GRIB.CMC2_reg_PRMSL_MSL_0_ps15km_2014030412_P009.Grb2_2014-03-04_150055
```

These files of Canadian model data have three date/time specifications: the directory contains the receipt time by the EDEX manual endpoint, the first time in the filename is the model run time followed by the forecast hour (e.g., 2014030412_P018 is the 18 hour forecast), and the last time is the receipt time by LDAD. Notice that this particular example has a model run time (12Z) that spans two hours of receipt times (14Z and 15Z).

Note: The exact filename structures of the Canadian (and other models that arrive via LDAD) may vary from WFO to WFO depending on the local LDAD configuration.

Step 2. Analyze existing patterns in RAW_DATA.xml. The entry in the baseline RAW_DATA.xml (in OB14.2.4) for hourly subdirectories in /data_store/manual/grib (❶) only handles files that contain LAPS and MSAS in their filenames. Thus, the baseline patterns don't apply to Canadian Model data.

Step 3. Modify RAW_DATA.xml to incorporate this data, and verify the modifications take effect using the Case Creation GUI.

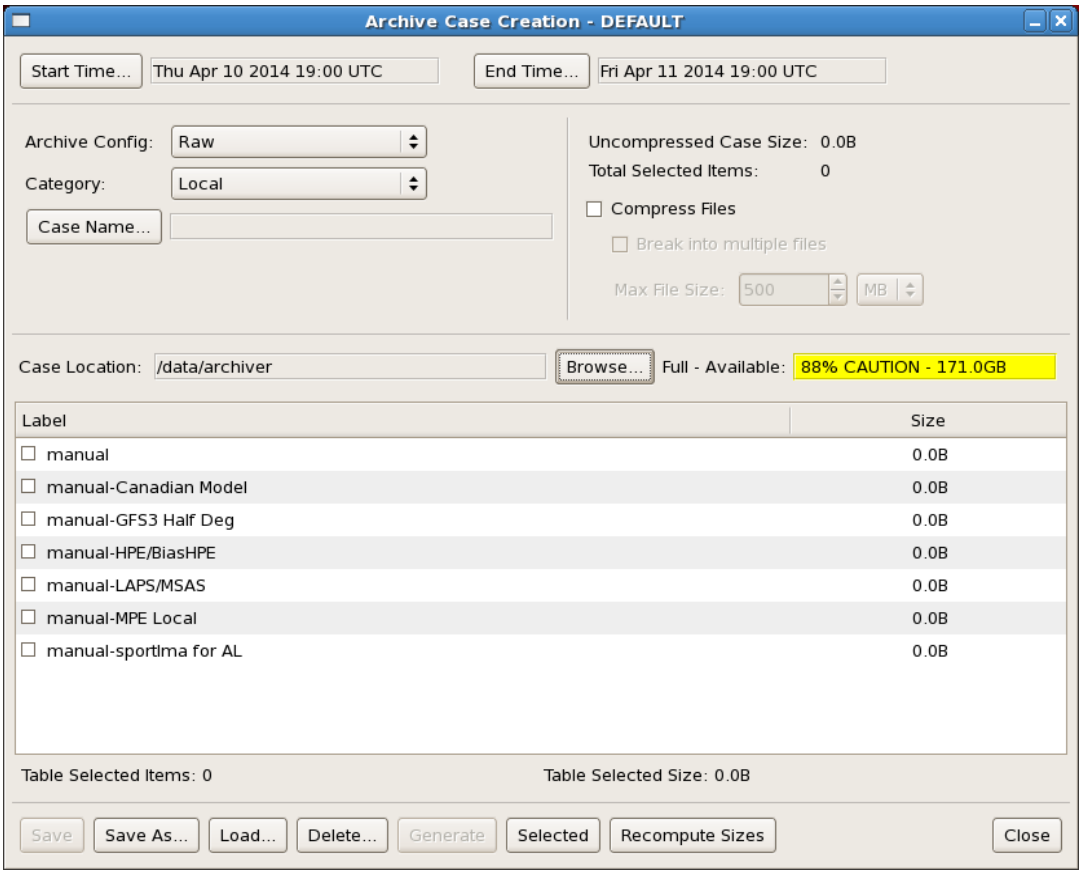
The Canadian model data need to use the date and time of the model run, which is the first date in the actual filename.

Distinguishing Characteristic: "CMC2" is a distinguishing characteristic of this data. The filenames also contain a domain identifier, so you could archive the northern hemisphere ("CMC2_nh") separately from the regional version if you wish ("CMC2_reg").

```
<dataSet>
   <dirPattern>(manual)/grib/\d{8}/\d{2}</dirPattern>
   <filePattern>.*(CMC2_).*(\d{4})(\d{2})(\d{2})(\d{2}).*</filePattern>
   <timeType>Date</timeType>
   <displayLabel>{1}-Canadian Model</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample file: manual/grib/20140304/16/LDAD-GRIB.CMC2_nh_HGT_ISBL_500_latlon.6x.6_2014030412_P000.Grb2_2014-03-04_160348

The result of adding the dataset entry for the Canadian Model to RAW_DATA.xml results in this Case Creation GUI:



Data Example 8. Raw Data: HPC Winter Weather Desk Products

Step 1. Analyze the file/directory structure. Here are a few files from the HPC Winter Weather Desk. All are located in /data_store/manual/grib.

```
manual/grib/20140304/06/LDAD-GRIB.2013-
2014_HPCWWD_2.5km_6h_zr_2014030412f072.Grb2_2014-03-04_061327
manual/grib/20140304/06/LDAD-GRIB.2013-
2014_HPCWWD_2.5km_6h_zr_2014030412f060.Grb2_2014-03-04_061341
manual/grib/20140304/06/LDAD-GRIB.2013-
2014_HPCWWD_2.5km_6h_snow_2014030412f066.Grb2_2014-03-04_061343
manual/grib/20140304/14/LDAD-GRIB.2013-2014_HPCWWD_2.5km_climo_snoden_spring-
MAM.Grb2_2014-03-04_140150
manual/grib/20140304/14/LDAD-GRIB.2013-
2014_HPCWWD_2.5km_6h_snow_2014030412f072.Grb2_2014-03-04_140230
manual/grib/20140304/16/LDAD-GRIB.2013-
2014_HPCWWD_2.5km_nam32_snoden_2014030500f006.Grb2_2014-03-04_160220
manual/grib/20140304/16/LDAD-GRIB.2013-
2014_HPCWWD_2.5km_gfs_snoden_2014030500f036.Grb2_2014-03-04_160230
manual/grib/20140304/16/LDAD-GRIB.2013-
```

```
2014_HPCWWD_2.5km_gfs_snoden_2014030500f042.Grb2_2014-03-04_160235
manual/grib/20140304/16/LDAD-GRIB.2013-
2014_HPCWWD_2.5km_nam32_snoden_2014030500f024.Grb2_2014-03-04_160241
```

The winter weather grids from the HPC Winter Weather Desk have two different structures. All files except for the climatological product (*climo_snowden_spring*) follow this pattern: **manual/grib/YYYYMMDD/HH/LDAD-GRIB.2013-2014_HPCWWD_2.5km_{model}_{product}_YYYYMMDDHHfFFF.Grb2_YYYY-MM-DD_HHMMSS.** The climatology does not have a model run time in the filename.

Step 2. Analyze existing patterns in RAW_DATA.xml. The discussion above for the Canadian model also applies to the HPC Winter Weather Desk grib products.

Step 3. Modify RAW_DATA.xml to incorporate this data, and verify the modifications take effect using the Case Creation GUI.

Two different dataset entries are needed in RAW_DATA.xml to handle both types of files. Most of the products need to use the model cycle time, which is the first date in the filename. Because the climatology product doesn't have a model time, we need a special entry for those files and must resort to the directory time.

Distinguishing Characteristic: "HPCWWD" is a distinguishing characteristic of this data. The climatological products are further identified by "climo".

Non-Climatological products:

```
<dataSet>
   <dirPattern>(manual)/grib/\d{8}/\d{2}</dirPattern>
   <filePattern>.*(HPCWWD).*_(\d{4})(\d{2})(\d{2})(\d{2}).*</filePattern>
   <timeType>Date</timeType>
   <displayLabel>{1}-2013-2014 HPCWWD</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample file:

```
manual/grib/20140304/04/LDAD-GRIB.2013-
2014_HPCWWD_2.5km_6h_snow_2014030412f018.Grb2_2014-03-04_044224
```

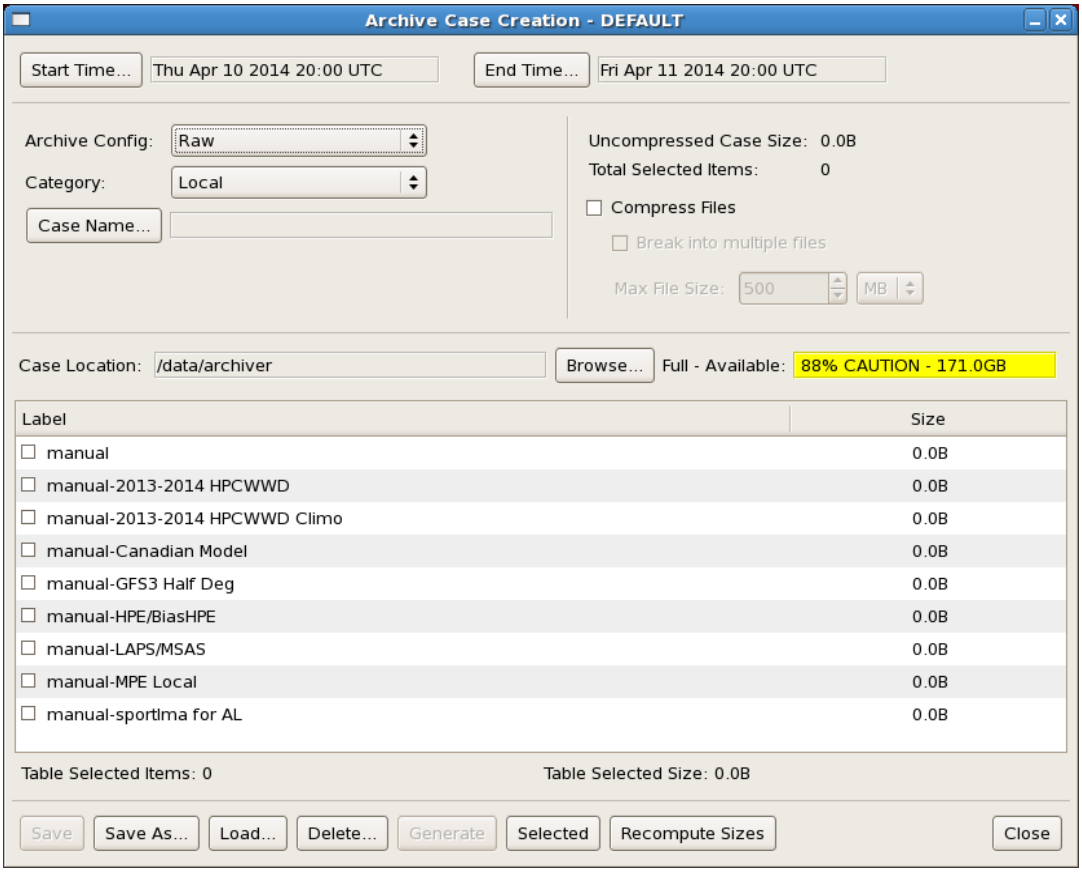
Climatological Product:

```
<dataSet>
   <dirPattern>(manual)/grib/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <filePattern>.*(HPCWWD).*_climo.*</filePattern>
   <timeType>Date</timeType>
   <displayLabel>{1}-2013-2014 HPCWWD Climo</displayLabel>
   <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

Sample file:

manual/grib/20140304/04/LDAD-GRIB.2013-2014_HPCWWD_2.5km_climo_snoden_spring-MAM.Grb2_2014-03-04_042347

The result of adding these two entries for HPCWWD products results in this Case Creation GUI:



Data Example 9. Raw Data: TAFs

Step 1. Analyze the file/directory structure. Here are a few TAF files that were issued by a local WFO.

```
manual/nctaf/20140327/05/003-KARXTAFRST-FTUS43-KARX-1403270537-AAA-1395898647
manual/nctaf/20140327/05/003-KARXTAFRST-FTUS43-KARX-1403270551-AAB-1395899510
manual/nctaf/20140327/05/004-KARXTAFLSE-FTUS43-KARX-1403270525-    -1395897900
manual/nctaf/20140327/05/004-KARXTAFRST-FTUS43-KARX-1403270525-    -1395897900
manual/nctaf/20140327/07/003-KARXTAFLSE-FTUS43-KARX-1403270706-AAA-1395904019
manual/nctaf/20140327/07/003-KARXTAFRST-FTUS43-KARX-1403270746-AAC-1395906398
manual/nctaf/20140327/08/003-KARXTAFLSE-FTUS43-KARX-1403270847-AAB-1395910039
manual/nctaf/20140327/09/003-KARXTAFRST-FTUS43-KARX-1403270907-AAD-1395911278
manual/nctaf/20140327/11/003-KARXTAFLSE-FTUS43-KARX-1403271125-    -1395919500
```

```
manual/nctaf/20140327/11/003-KARXTAFRST-FTUS43-KARX-1403271125-    -1395919500
```

Note that some of the filenames contain spaces. There are also three date/time specifications in these files. The date/time that is part of the directory structure is the date/time the TAF was received by the manual endpoint to be ingested by EDEX. The first ten-digit number ("1403…") is the issuance time expressed in YYMMDDHHMM format. The last ten-digit number is the issuance time expressed in Epoch seconds (the number of seconds since 1/1/1970). RAW_DATA.xml could utilize the Epoch time to identify these files, but we can also use the directory date/time for performance.

Step 2. Analyze existing patterns in RAW_DATA.xml. There is an entry (❹) in the baseline RAW_DATA.xml in OB14.2.4 that was designed to archive TAF files. This entry captures any file, containing spaces or not, that is located in an hourly subdirectory of /data_store/manual/taf or /data_store/manual/nctaf. There is no need for further modifications, so long as the directory date/hour is sufficient to capture the issuance time of the TAF.

Data Example 10. Raw Data: SHEF (Standard Hydrometeorological Exchange Format) Products

Step 1. Analyze the file/directory structure. Here are a few SHEF files from /data_store/manual/shef.

```
manual/shef/20140304/13/MKERR8ARX
manual/shef/20140304/13/MKERR8ARX1393938335
manual/shef/20140304/13/RTPARX.wan1393938891
manual/shef/20140304/14/MKERR8ARX
manual/shef/20140304/14/MKERR8ARX1393941930
manual/shef/20140304/14/RTPARX.wan1393942761
manual/shef/20140304/15/MKERR8ARX
manual/shef/20140304/15/MKERR8ARX1393945533
manual/shef/20140304/15/RTPARX.wan1393945697
manual/shef/20140304/15/LKZM5
manual/shef/20140304/16/MKERR8ARX1393949137
```

Some of the SHEF data have Epoch seconds appended at the end of the filename. But some have no date/time identification, other than the directory.

Step 2. Analyze existing patterns in RAW_DATA.xml. The baseline RAW_DATA.xml file in OB14.2.4 includes an entry (❺) designed specifically to pick up all files in hourly subdirectories of /data_store/manual/shef. There is no need for further configuration for locally-saved SHEF files.

Data Example 11. Raw Data: Local Text Products

Step 1. Analyze the file/directory structure. The files in /data_store/manual/text consist of text products issued by the local WFO. Note that some of the products listed below were issued by the ARX forecast office while they were in service backup for the MPX WFO. All of the text products have the

issuance time represented by Epoch seconds (ten digits that describe the number of seconds elapsed since midnight on 1/1/1970; for example, 1395867146) appended at the end of the filename.

```
manual/text/20140326/20/KARXAFDARX-0-1395867146
manual/text/20140326/20/KMPXAFDMPX-0-1395865297
manual/text/20140326/20/KMPXZFPMPX-1-1395866235
manual/text/20140326/20/KMPXZFPMPX-3-1395866893
manual/text/20140326/20/RWSARX.wan1395865506
manual/text/20140327/02/000-KARXVFTARX-NXUS98-KARX-1403270250-___-1395888650
manual/text/20140327/04/KARXAFDARX-2-1395895425
manual/text/20140327/05/NOWARX.wan1395897267
manual/text/20140327/08/000-KARXVFTARX-NXUS98-KARX-1403270850-___-1395910250
manual/text/20140327/08/KARXAFDARX-0-1395909608
manual/text/20140327/08/KARXNOWARX-1-1395909893
manual/text/20140327/09/KARXHWOARX-2-1395911228
manual/text/20140327/10/PNSARX.wan1395916182
manual/text/20140327/11/AFDARX.wan1395919462
manual/text/20140327/12/KARXNOWARX-3-1395922291
manual/text/20140327/14/000-KARXVFTARX-NXUS98-KARX-1403271450-___-1395931851
manual/text/20140327/15/RWSARX.wan1395934251
manual/text/20140327/19/KARXFWFARX-3-1395949502
manual/text/20140327/19/KARXRECARX-2-1395947403
```

Step 2. Analyze existing patterns in RAW_DATA.xml. There is an existing dataSet entry (❻) in RAW_DATA.xml intended for text products. This entry expects the products to be in /data_store/manual/text/YYYYDDMM/HH, but uses the Epoch seconds for the date/time of the file. So long as the filenames in /data_store/manual/text/YYYYMMDD/HH all end with ten digits describing Epoch seconds, there should be no need for further configuration.

Data Example 12. Raw Data: Local Warnings

Step 1. Analyze the file/directory structure. Here are a few warning products that were issued by a local WFO.

```
warning/20140228/15/KARXWSWARX-0-1393599884
warning/20140228/23/KARXSPSARX-1-1393628598
warning/20140228/23/KARXSPSARX-2-1393628786
warning/20140303/14/KARXWSWARX-0-1393858642
warning/20140327/21/SPSARX.wan1395957419
warning/20140327/22/SPSARX.wan1395957732
warning/20140328/01/FLSARX.wan1395970260
warning/20140328/03/FLSARX.wan1395975686
warning/20140328/05/FLSARX.wan1395983461
warning/20140328/05/FLSARX.wan1395983893
```

The warnings follow a similar pattern as the text products in the section above. Regardless of the software application (GFE or WarnGen) that produced the warnings, the issuance time is appended at the end of the filenames, in Epoch seconds format.

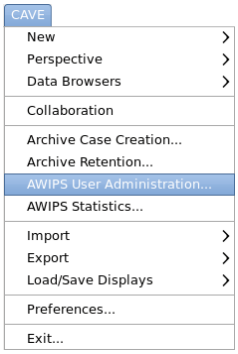
Step 2. Analyze existing patterns in RAW_DATA.xml. The discussion above for text products also applies to warnings, because there is a similar entry (❼) in RAW_DATA.xml for warnings. There should be no need for further configuration changes to archive warnings.

Appendix 1: Adding Archive Permissions Using the AWIPS User Admin GUI

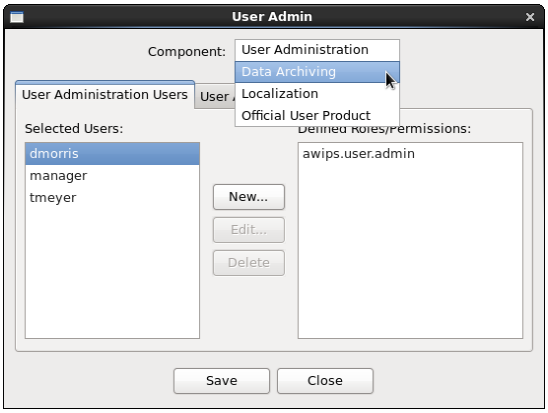
In order to use the AWIPS-2 Archiver Retention and Case Creation GUIs and to perform the steps in this Exercise, several permissions must be granted to one or more users using the AWIPS User Administration GUI. These steps assume your user has already been added to your SITE's version of awipsUserAdminRoles.xml in common_static/site/<SITE>/roles, so you can access the AWIPS User Admin GUI in the first place.

Task 1. Add the Archive Retention and Archive Case Creation permissions.

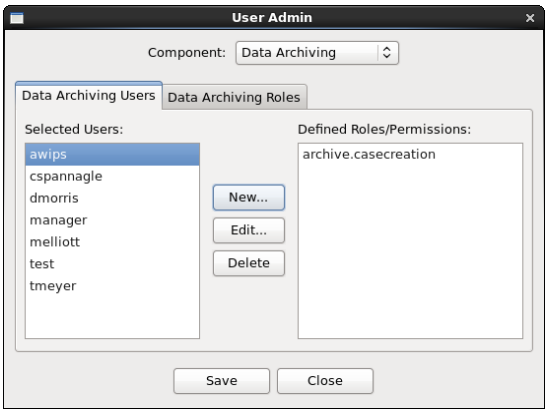
1. Start the AWIPS User Administration GUI (CAVE ► AWIPS User Administration).

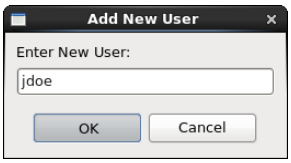


2. Change the Component to Data Archiving.

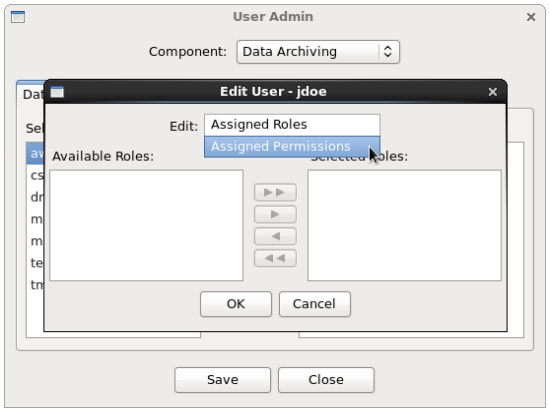


3. Select a user from the list. If you want to give permission to a user that does not exist in the list, click the "New" button and enter the new user to the list. In our case, we'll add jdoe as a new user.

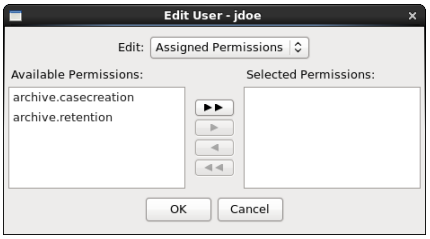




4. In the Edit User dialog, change the Edit dropdown box from “Assigned Roles” to “Assigned Permissions”.



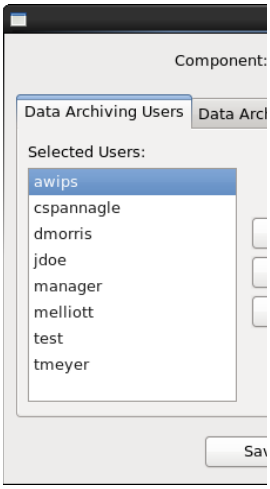
5. Click on the double right arrow button to move both permissions (case creation and retention) to the “Selected Permissions” list.



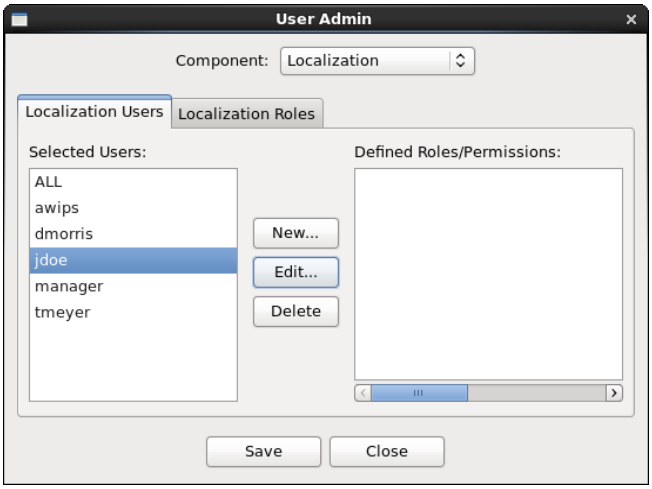
6. Click OK.

Task 2. Add the com.raytheon.localization.site/common_static/archiver/purger permission in order to use the localization perspective to edit archiver configuration files (RAW_DATA.xml and PROCESSED_DATA.xml).

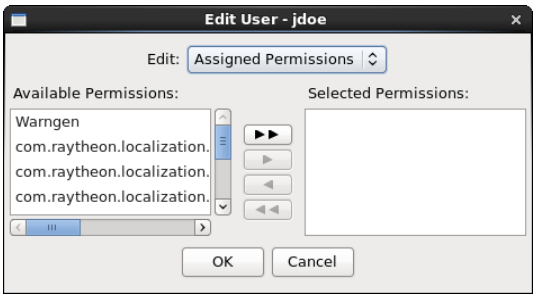
1. Change the Component to Localization.



2. Select a user from the list. If the user you want to give the permission to does not exist in the list, click the "New" button and enter the new user to the list. In the case, we'll use the jdoe our user that was added in Task 1.



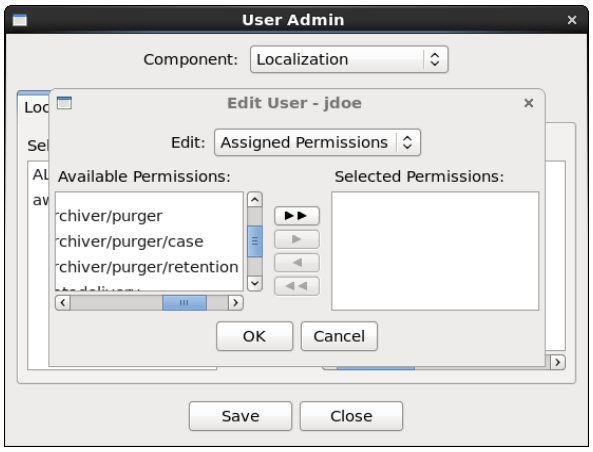
3. In the Edit User dialog, change Edit from “Assigned Roles” to “Assigned Permissions”.

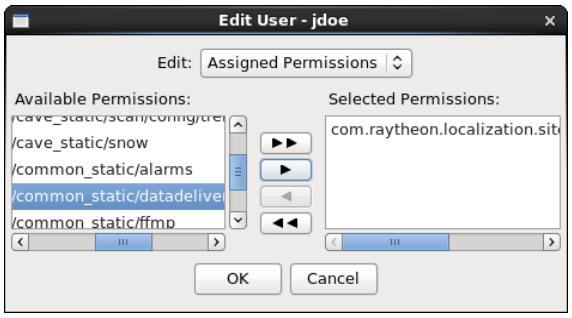


4. Select the archiver permissions. There are actually three to select:
   com.raytheon.localization.site/common_static/archiver/purger
   com.raytheon.localization.site/common_static/archiver/purger/case
   com.raytheon.localization.site/common_static/archiver/purger/retention

   They are probably about half-way down the list.

5. Move each permission in succession to the selected permissions window.





6. Click OK.
7. Click Save.

Appendix 2: Additional Examples

This appendix contains additional examples of archivable files and corresponding recommendations for entries in either RAW_DATA.xml or PROCESSED_DATA.xml. The examples in this appendix were collected from a variety of WFOs, and represent a cross-section of new or local types of data that may be encountered across the NWS. They also represent a variety of file-naming and storage conventions.

The appendix contains 18 additional examples:

1. Raw Data: LDAD Data
2. Raw Data: McIDAS Satellite
3. Raw Data: regionalsat
4. Raw Data: regionalsat from WFO HUN
5. Processed Data: regionalsat
6. Raw Data: manual/obs
7. Raw Data: SHEF from WFO HUN
8. Raw Data: unknown
9. Raw Data: /data_store/manual/grib from WFO HUN (GOES-R CI and NASA SPoRT LIS)
10. Raw Data: Ensembles from narre
11. Raw Data: WRF Model from Alaska Region
12. Raw Data: SPC Lightning for Alaska Region
13. Raw Data: HPC QPF NDFD grids
14. Raw Data: HRRR Model
15. Raw Data: NAM80 Model (beyond 60 hours)
16. Raw Data: University of Wisconsin Cloud-Top Cooling/Convective Initiation Products
17. Raw Data: Local WRF model at OUN
18. Rata Data: Multi-radar multi-sensor (MRMS) at OUN

Additional Example 1. Raw Data: LDAD Data

These files are from /data_store/ldad.

```
ldad/LDAD.mesonet.1398031585.FSLAWS_WI.decoded.xml
ldad/LDAD.mesonet.1398031851.widotmet.decoded.xml
ldad/LDAD.mesonet.NOS-NWLON.1398106800.06.msas_qc.decoded.1.xml
ldad/SXUS44 KWOHshef.1398031863
ldad/SXUS44 KWOHshef.1398031863.1
ldad/SXUS44 KWOHshef.1398031863.2
ldad/SXUS44 KWOHshef.1398031863.3
ldad/LDAD.mesonet.1398031851.NOS-NWLON.decoded.xml
ldad/LDAD.mesonet.1398031871.RAWS.decoded.xml
ldad/LDAD.mesonet.1398031871.NOS-NWLON2.decoded.xml
ldad/LDAD.mesonet.NOS-NWLON.1398027600.11.msas_qc.decoded.1.xml
ldad/LDAD.mesonet.NOS-NWLON.1398020400.06.msas_qc.decoded.xml
ldad/LDAD.mesonet.1398031911.FSLAWS_WI.decoded.xml
ldad/LDAD.mesonet.1398039041.widotmet.decoded.xml
ldad/LDAD.mesonet.1398039061.RAWS.decoded.xml
ldad/LDAD.mesonet.1398039081.NOS-NWLON2.decoded.xml
ldad/LDAD.hydro.1398039201.MMSD60.decoded.xml
```

```
ldad/LDAD.hydro.1398039441.CoCoRaHS.decoded.xml
ldad/LDAD.mesonet.1398039461.WIDNR_PRECIP.decoded.xml
```

All filenames in /data_store/ldad from this WFO contain the Epoch seconds representing the file time produced by LDAD. The Epoch seconds are immediately preceded by a dot (“.”) character. Note that the files for MSAS QC (*msas_qc.decoded*.xml) have the Epoch seconds truncated to the nearest 100 seconds. Closer inspection, not shown here, reveals that the next two digits contain the minutes of the file write time. There is not an easy way to handle this, so the sample entry below just treats these Epoch seconds as if they were not truncated in order to keep a simple, unified entry for all the files in this directory. An alternative method would be to use the file write time for the MSAS QC files. Note that there is no YYYYMMDD/HH directory structure, since this directory is not managed by the manual endpoint or created by the LDM’s pqact.conf file for the SBN.

No prior entry exists in the baseline RAW_DATA.xml for /data_store/ldad. This entry should probably be placed inside the Local category.

```
<dataSet>
    <dirPattern>(ldad)</dirPattern>
    <filePattern>.*\.(\d{10}).*</filePattern>
    <displayLabel>{1}</displayLabel>
    <timeType>EpochSec</timeType>
    <dateGroupIndices>2</dateGroupIndices>
</dataSet>
```

Sample Files:

```
ldad/SXUS44 KWOHshef.1398031851
ldad/SXUS44 KWOHshef.1398031851.1
ldad/LDAD.mesonet.1398031585.FSLAWS_WI.decoded.xml
ldad/LDAD.mesonet.NOS-NWLON.1398106800.06.msas_qc.decoded.1.xml
ldad/LDAD.hydro.1398039201.MMSD60.decoded.xml
```

Additional Example 2. Raw Data: McIDAS Satellite

Raw data for the McIDAS satellite plugin are stored in the /data_store/manual/satellite.mcidas directory. Here is a listing of a sampling of two hours of data from this directory from one WFO. McIDAS files from other WFOs may have somewhat different structures, depending on the LDAD configuration and the source of the McIDAS files. This WFO obtained these files from CIRA (Cooperative Institute for Research in the Atmosphere).

```
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1830_2014-04-20_191452
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_LCFOG_20140420_1900_2014-04-20_191900
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1845_2014-04-20_192413
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_EC_LCFOG_20140420_1915_2014-04-20_192437
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_EC_GEOC_20140420_1845_2014-04-20_192438
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_RGB_AIRMASS_20140420_1901_2014-04-
20_194041
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1900_2014-04-20_194422
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_LCFOG_20140420_1930_2014-04-20_194838
```

```
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_EC_LCFOG_20140420_1945_2014-04-20_195430
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_HI_ORI_20140420_2100_2014-04-20_195440
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_EC_GEOC_20140420_1915_2014-04-20_195441
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1915_2014-04-20_201410
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_WC_LCFOG_20140420_2000_2014-04-20_201852
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_EC_LCFOG_20140420_2015_2014-04-20_202432
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1930_2014-04-20_203444
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_EC_GEOC_20140420_1945_2014-04-20_203504
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_RGB_AIRMASS_20140420_2001_2014-04-
20_204034
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1945_2014-04-20_204255
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_WC_LCFOG_20140420_2030_2014-04-20_204821
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_HI_ORI_20140421_0000_2014-04-20_205427
satellite.mcidas/20140420/20/LDAD-MCIDAS-CIRA_AWIPS2_EC_GEOC_20140420_2015_2014-04-20_205939
```

These filenames contain three times (the directory, the actual data time and the LDAD time). The data time is in YYYYMMDD_HHMM format. In addition, a time lag is apparent because there are files in the 20140420/20 directory labeled 1930 and 1945 UTC. Because of these time differences, the entry below uses the date/time that represents the data time.

No entry in the baseline RAW_DATA.xml file applies to satellite.mcidas because no <dirPattern> tag exists for satellite.mcidas. The satellite.mcidas entry below was added to the Satellite category, so that category is shown in its entirety:

```
<category>
   <name>Satellite</name>
   <extRetentionHours>168</extRetentionHours>
   <dataSet>
      <dirPattern>sat/(\d{4})(\d{2})(\d{2})/(\d{2})/(.*)</dirPattern>
      <displayLabel>{5}</displayLabel>
      <dateGroupIndices>1,2,3,4</dateGroupIndices>
   </dataSet>
   <dataSet>
      <dirPattern>manual/(satellite.mcidas)/.*/.*</dirPattern>
      <filePattern>.*_(\d{4})(\d{2})(\d{2})_(\d{2})(\d{2})_.*</filePattern>
      <displayLabel>{1}</displayLabel>
      <dateGroupIndices>2,3,4,5</dateGroupIndices>
   </dataSet>
</category>
```

Sample Files:

```
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1830_2014-04-20_191452
satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_LCFOG_20140420_1900_2014-04-20_191900
```

Additional Example 3. Raw Data: regionalsat

Some satellite products are ingested using the regionalsat plugin. This plugin stores its processed data alongside the data from the regular satellite plugin. See Additional Example 5 below for details on how to handle the processed data.

The raw data are in /data_store/manual/regionalsat, the processed data are in /awips2/edex/data/hdf5/satellite, and the archived versions are in /archive/satellite. Here are some sample raw data files for a two-hour period:

```
regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_MVFRPROB_20140421_1345.55779862_2014-04-21_140125

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_IFRPROB_20140421_1345.55779865_2014-04-21_140145

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_LIFRPROB_20140421_1345.55779712_2014-04-21_140225

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_IFRPROB_20140421_1400.55779865_2014-04-21_141802

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_MVFRPROB_20140421_1415.55779862_2014-04-21_143128

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_IFRPROB_20140421_1415.55779865_2014-04-21_143148

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_LIFRPROB_20140421_1415.55779712_2014-04-21_143228

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOE_CONUS_4KM_IFRPROB_20140421_1415.55779870_2014-04-21_143715

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOE_CONUS_4KM_MVFRPROB_20140421_1415.55779868_2014-04-21_143715

regionalsat/20140421/14/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOE_CONUS_4KM_LIFRPROB_20140421_1415.55779711_2014-04-21_143740

regionalsat/20140421/15/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_MVFRPROB_20140421_1445.55779862_2014-04-21_150108

regionalsat/20140421/15/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOW_CONUS_4KM_IFRPROB_20140421_1445.55779865_2014-04-21_150133

regionalsat/20140421/15/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOE_CONUS_4KM_LIFRPROB_20140421_1445.55779711_2014-04-21_151930

regionalsat/20140421/15/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOE_CONUS_4KM_IFRPROB_20140421_1445.55779870_2014-04-21_151930

regionalsat/20140421/15/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOE_CONUS_4KM_MVFRPROB_20140421_1445.55779868_2014-04-21_151930

regionalsat/20140421/15/LDAD-REGIONALSAT-
SSEC_AWIPS_MODIS_EAST_4KM_TPW_TERRA_20140421_1444.7360_2014-04-21_152410

regionalsat/20140421/15/LDAD-REGIONALSAT-
SSEC_AWIPS_MODIS_EAST_4KM_LI_TERRA_20140421_1444.7363_2014-04-21_152430

regionalsat/20140421/15/LDAD-REGIONALSAT-
SSEC_AWIPS_MODIS_EAST_4KM_TT_TERRA_20140421_1444.7364_2014-04-21_152450

regionalsat/20140421/15/LDAD-REGIONALSAT-
SSEC_AWIPS_MODIS_EAST_4KM_KI_TERRA_20140421_1444.7365_2014-04-21_152450
```

The product times are in the filenames and they do not necessarily match the directory time. Files representing 1345 UTC are in the 20140421/14 directory. The regionalsat <dirPattern> can be combined with satellite.mcidas from Additional Example 2, if the satellite.mcidas <filePattern> is tweaked to remove the underscore after the date (this is the date represented by the \d{4} and \d{2} expressions in the <filePattern>; this works because the .* after the underscore can pick up that underscore character in the McIDAS files and because the dot character represents any single character).

```
<category>
    <name>Satellite</name>
    <extRetentionHours>168</extRetentionHours>
    <dataSet>
        <dirPattern>sat/(\d{4})(\d{2})(\d{2})/(\d{2})/(.*)</dirPattern>
        <displayLabel>{5}</displayLabel>
        <dateGroupIndices>1,2,3,4</dateGroupIndices>
    </dataSet>
    <dataSet>
        <dirPattern>manual/(satellite.mcidas|regionalsat)/.*/.*</dirPattern>
        <filePattern>.*_(\d{4})(\d{2})(\d{2})_(\d{2})(\d{2}).*</filePattern>
        <displayLabel>{1}</displayLabel>
        <dateGroupIndices>2,3,4,5</dateGroupIndices>
    </dataSet>
</category>
```

Sample Files:

```
regionalsat/20140421/15/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOE_CONUS_4KM_MVFRPROB_20140421_1445.55779868_2014-04-21_151930

regionalsat/20140421/15/LDAD-REGIONALSAT-
SSEC_AWIPS_MODIS_EAST_4KM_TPW_TERRA_20140421_1444.7360_2014-04-21_152410

satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1830_2014-04-20_191452

satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_LCFOG_20140420_1900_2014-04-20_191900
```

Additional Example 4. Raw Data: regionalsat from WFO HUN

The Huntsville WFO receives also processes regionalsat data, but they receive their data from NASA SPoRT.

```
regionalsat/20140421/19/20140421_1754_sport_modis_sre_vis
regionalsat/20140421/19/20140421_1800_sport_hybrid_sre_wvairmass
regionalsat/20140421/19/20140421_1755_sport_modis_sre_dust
regionalsat/20140421/19/20140421_1754_sport_modis_sre_dust
regionalsat/20140421/19/20140421_1800_sport_hybrid_srw_wvairmass
regionalsat/20140421/19/20140421_1755_sport_modis_sre_color
regionalsat/20140421/19/20140421_1755_sport_modis_srw_color
regionalsat/20140421/19/20140421_1800_sport_hybrid_sre_fogNtmicro
regionalsat/20140421/19/20140421_1754_sport_modis_sre_color
regionalsat/20140421/19/20140421_1845_sport_hybrid_srw_11um
```

```
regionalsat/20140421/19/20140421_1755_sport_modis_sre_vis
regionalsat/20140421/19/20140421_1845_sport_hybrid_srw_39um
regionalsat/20140421/19/20140421_1845_sport_hybrid_sre_11um
regionalsat/20140421/19/20140421_1845_sport_hybrid_srw_wvairmass
regionalsat/20140421/19/20140421_1845_sport_hybrid_sre_39um
regionalsat/20140421/19/20140421_1755_sport_modis_srw_vis
regionalsat/20140421/19/20140421_1800_sport_hybrid_srw_fogNtmicro
regionalsat/20140421/19/20140421_1755_sport_modis_sre_false
regionalsat/20140421/19/20140421_1755_sport_modis_srw_false
regionalsat/20140421/19/20140421_1845_sport_hybrid_sre_wvairmass
regionalsat/20140421/19/20140421_1902_sport_hybrid_srw_11um
regionalsat/20140421/19/20140421_1902_sport_hybrid_srw_wvairmass
regionalsat/20140421/19/20140421_1902_sport_hybrid_sre_11um
regionalsat/20140421/19/20140421_1902_sport_hybrid_srw_39um
regionalsat/20140421/19/20140421_1902_sport_hybrid_sre_39um
regionalsat/20140421/19/20140421_1754_sport_modis_sre_false
regionalsat/20140421/19/20140421_1902_sport_hybrid_sre_wvairmass
regionalsat/20140421/19/20140421_1901_sport_goesSndr_conus_airmass
regionalsat/20140421/19/20140421_1915_sport_hybrid_srw_11um
regionalsat/20140421/19/20140421_1915_sport_hybrid_srw_wvairmass
regionalsat/20140421/19/20140421_1915_sport_hybrid_sre_11um
regionalsat/20140421/19/20140421_1915_sport_hybrid_srw_39um
regionalsat/20140421/19/20140421_1915_sport_hybrid_sre_39um
regionalsat/20140421/19/20140421_1915_sport_hybrid_sre_wvairmass
regionalsat/20140421/19/20140421_1932_sport_hybrid_srw_11um
regionalsat/20140421/19/20140421_1932_sport_hybrid_srw_wvairmass
regionalsat/20140421/19/20140421_1932_sport_hybrid_srw_39um
regionalsat/20140421/19/20140421_1932_sport_hybrid_sre_11um
regionalsat/20140421/19/20140421_1932_sport_hybrid_sre_39um
```

Although the regionalsat filenames from Additional Examples 3 and 4 may appear at first glance to have radically different conventions, the date format for both kinds of files is YYYYMMDD_HHMM. The regionalsat entry from Additional Example 3 can be used for HUN if the underscore prior to the date in the <filePattern> is removed (this underscore is the first underscore character in the <filePattern> in Additional Example 3). This works because the .* expression can match zero or more characters. You may find it helpful to carefully compare the colorized sample files below with those from the previous examples.

```
<category>
   <name>Satellite</name>
   <extRetentionHours>168</extRetentionHours>
   <dataSet>
       <dirPattern>sat/(\d{4})(\d{2})(\d{2})/(\d{2})/(.*)</dirPattern>
       <displayLabel>{5}</displayLabel>
       <dateGroupIndices>1,2,3,4</dateGroupIndices>
   </dataSet>
   <dataSet>
       <dirPattern>manual/(satellite.mcidas|regionalsat)/.*/.*</dirPattern>
       <filePattern>.*(\d{4})(\d{2})(\d{2})_(\d{2})(\d{2}).*</filePattern>
       <displayLabel>{1}</displayLabel>
       <dateGroupIndices>2,3,4,5</dateGroupIndices>
   </dataSet>
</category>
```

Sample Files:

```
regionalsat/20140421/19/20140421_1932_sport_hybrid_sre_wvairmass

regionalsat/20140421/15/LDAD-REGIONALSAT-SSEC_AWIPS_GEOCAT-
GEOE_CONUS_4KM_MVFRPROB_20140421_1445.55779868_2014-04-21_151930

regionalsat/20140421/15/LDAD-REGIONALSAT-
SSEC_AWIPS_MODIS_EAST_4KM_TPW_TERRA_20140421_1444.7360_2014-04-21_152410

satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_GEOC_20140420_1830_2014-04-20_191452

satellite.mcidas/20140420/19/LDAD-MCIDAS-CIRA_AWIPS2_WC_LCFOG_20140420_1900_2014-04-20_191900
```

Additional Example 5. Processed Data: regionalsat

As stated in Additional Example 3, the processed data for the regionalsat plugin (and for satellite.mcidas, as well) is actually stored in the same directory as the regular, baseline satellite plugin. This example illustrates why it is necessary to check existing patterns in RAW_DATA.xml and PROCESSED_DATA.xml because the entry designed for the baseline plugin may actually already archive the non-baseline data, or the baseline entry may need a slight tweak.

Here is a directory listing of /archive /satellite from a WFO that processes satellite data using the baseline satellite plugin, satellite.mcidas, and regionalsat. The non-baseline directories are indicated in blue.

```
Alaska National     AREA2402                              MODIS Sea Sfc Temperature Sum 1km (F)
Alaska Regional     conusOne                              MODIS Total Totals 4km (C)
AREA0133            East CONUS                            NH Composite - Meteosat-GOES E-GOES W-GMS
AREA0988            eastConus                             Northern Hemisphere Composite
AREA0998            Hawaii National                       Puerto Rico National
AREA1881            Hawaii Regional                       Puerto Rico Regional
AREA1891            MODIS 11um - 3.7um Product 1km (C)    Supernational
AREA2201            MODIS K Index 4km (C)                 West CONUS
AREA2202            MODIS Land Sfc Temperature 1km (F)    westConus
AREA2221            MODIS Land Sfc Temperature Sum 1km (F)
AREA2222            MODIS Lifted Index 4km (C)
```

The AREA directories come from satellite.mcidas and the additional conusOne, eastConus, westConus, and MODIS directories may come from either satellite.mcidas or regionalsat. Because the baseline directories are known to archive properly, it's necessary to check the new directories against the baseline ones to see if there are any differences in structure. Here is a listing of the East CONUS baseline directory (performed by executing an "`ls E*/*`" command from /archive/satellite). Note that hourly directories exist in each sensor's directory. These hourly directories contain the .h5 and .bin.* files (not shown for clarity and to save space).

```
East CONUS/Imager 11 micron IR:
satellite-2014-04-21-01    satellite-2014-04-21-02    satellite-2014-04-21-03    satellite-2014-04-21-04
```

```
East CONUS/Imager 13 micron (IR):
satellite-2014-04-21-01   satellite-2014-04-21-02   satellite-2014-04-21-03   satellite-2014-04-21-04

East CONUS/Imager 6.7-6.5 micron IR (WV):
satellite-2014-04-21-01   satellite-2014-04-21-02   satellite-2014-04-21-03   satellite-2014-04-21-04

East CONUS/Imager Visible:
satellite-2014-04-21-00   satellite-2014-04-21-01   satellite-2014-04-21-02   satellite-2014-04-21-03

East CONUS/Low cloud base imagery:
satellite-2014-04-21-01   satellite-2014-04-21-02   satellite-2014-04-21-03   satellite-2014-04-21-04

East CONUS/Sounder 11.03 micron imagery:
satellite-2014-04-21-00   satellite-2014-04-21-01   satellite-2014-04-21-02   satellite-2014-04-21-03

East CONUS/Sounder 14.06 micron imagery:
satellite-2014-04-21-00   satellite-2014-04-21-01   satellite-2014-04-21-02   satellite-2014-04-21-03

East CONUS/Sounder Visible imagery:
satellite-2014-04-21-00   satellite-2014-04-21-01   satellite-2014-04-21-02   satellite-2014-04-21-01
```

Here is a listing of some the MODIS directories using a similar command ("`ls M*/*`"). The MODIS directory structure contains one less level of directories. For example, East CONUS *.bin and *.h5 files are 5 directories deep (e.g., archive, satellite, East CONUS, Imager Visible, satellite-2014-04-01-00) from /archive versus 4 directories deep for MODIS as shown in these directory listings. This means that the baseline PROCESSED_DATA.xml entry likely will not work for MODIS because it is searching for a level of directories that do not exist.

```
MODIS 11um - 3.7um Product 1km (C)/satellite-2014-04-21-02:
satellite-2014-04-21-02.bin.1   satellite-2014-04-21-02.h5

MODIS 11um - 3.7um Product 1km (C)/satellite-2014-04-21-04:
satellite-2014-04-21-04.bin.1   satellite-2014-04-21-04.h5

MODIS K Index 4km (C)/satellite-2014-04-21-02:
satellite-2014-04-21-02.bin.1   satellite-2014-04-21-02.h5

MODIS K Index 4km (C)/satellite-2014-04-21-04:
satellite-2014-04-21-04.bin.1   satellite-2014-04-21-04.h5

MODIS Land Sfc Temperature 1km (F)/satellite-2014-04-21-02:
satellite-2014-04-21-02.bin.1   satellite-2014-04-21-02.h5

MODIS Land Sfc Temperature 1km (F)/satellite-2014-04-21-04:
satellite-2014-04-21-04.bin.1   satellite-2014-04-21-04.h5

MODIS Land Sfc Temperature Sum 1km (F)/satellite-2014-04-21-02:
satellite-2014-04-21-02.bin.1   satellite-2014-04-21-02.h5

MODIS Land Sfc Temperature Sum 1km (F)/satellite-2014-04-21-04:
satellite-2014-04-21-04.bin.1   satellite-2014-04-21-04.h5

MODIS Lifted Index 4km (C)/satellite-2014-04-21-02:
satellite-2014-04-21-02.bin.1   satellite-2014-04-21-02.h5

MODIS Lifted Index 4km (C)/satellite-2014-04-21-04:
satellite-2014-04-21-04.bin.1   satellite-2014-04-21-04.h5
```

```
MODIS Sea Sfc Temperature Sum 1km (F)/satellite-2014-04-21-02:
satellite-2014-04-21-02.bin.1   satellite-2014-04-21-02.h5

MODIS Sea Sfc Temperature Sum 1km (F)/satellite-2014-04-21-04:
satellite-2014-04-21-04.bin.1   satellite-2014-04-21-04.h5

MODIS Total Totals 4km (C)/satellite-2014-04-21-02:
satellite-2014-04-21-02.bin.1   satellite-2014-04-21-02.h5

MODIS Total Totals 4km (C)/satellite-2014-04-21-04:
satellite-2014-04-21-04.bin.1   satellite-2014-04-21-04.h5
```

Here is a listing of a few AREA* directories using a similar command ("`ls AREA*/*`"). The processed versions of McIDAS AREA files have a structure similar to the baseline satellite files, so below we only need to work on a configuration to handle the MODIS files.

```
AREA0988/CIRA-LCFOG:
satellite-2014-04-21-01  satellite-2014-04-21-02  satellite-2014-04-21-03  satellite-2014-04-21-04

AREA0998/CIRA-LCFOG:
satellite-2014-04-21-01  satellite-2014-04-21-02  satellite-2014-04-21-03  satellite-2014-04-21-04

AREA1881/Imager CIRA GeoColor:
satellite-2014-04-21-01  satellite-2014-04-21-02  satellite-2014-04-21-03  satellite-2014-04-21-04

AREA1891/Imager CIRA GeoColor:
satellite-2014-04-21-01  satellite-2014-04-21-02  satellite-2014-04-21-03  satellite-2014-04-21-04

AREA2201/CIRA-SIMWRF-IR:
satellite-2014-04-21-01  satellite-2014-04-21-02  satellite-2014-04-21-03  satellite-2014-04-21-04

AREA2202/CIRA-SIMWRF-WV:
satellite-2014-04-21-01  satellite-2014-04-21-02  satellite-2014-04-21-03  satellite-2014-04-21-04
```

Here is the satellite entry from the BASE version of PROCESSED_DATA.xml, along with a couple of sample files.

```
        <dataSet>
            <dirPattern>satellite/(.*)/(.*)/.*-(\d{4})-(\d{2})-(\d{2})-(\d{2})</dirPattern>
            <dateGroupIndices>3,4,5,6</dateGroupIndices>
            <displayLabel>{1}</displayLabel>
        </dataSet>
```

```
satellite/East CONUS/Imager Visible/satellite-2014-04-21-03/satellite-2014-04-21-03.bin.1
satellite/East CONUS/Imager Visible/satellite-2014-04-21-03/satellite-2014-04-21-03.h5
```

One dataset entry can contain multiple <dirPattern> tags. In this case we can modify the baseline satellite entry to add a <dirPattern> for MODIS (the first <dirPattern> below). However, when we do that, we also need to modify the other <dirPattern> to remove the parentheses around the second .* expression (shown in bright green above). This has to be done so that the regular expression groups represent the same information in both <dirPattern>s since there is only one <dateGroupIndices> tag that has to apply to both <dirPattern> entries. We can safely do that because Group 2 was not used

elsewhere in the entry (in either <dateGroupIndices> or <displayLabel>). The result is the .* shown in bold red and underlined below, and the groups for the <dateGroupIndices> (year/month/day/hour) agree between the two <dirPattern> tags. Finally, we can modify <dateGroupIndices> from 3,4,5,6 to 2,3,4,5, as shown below.

```
<dataSet>
    <dirPattern>satellite/(.*)/.*-(\d{4})-(\d{2})-(\d{2})-(\d{2})</dirPattern>
    <dirPattern>satellite/(.*)/.*/.*-(\d{4})-(\d{2})-(\d{2})-(\d{2})</dirPattern>
    <dateGroupIndices>2,3,4,5</dateGroupIndices>
    <displayLabel>{1}</displayLabel>
</dataSet>
```

Sample Files:

```
satellite/MODIS 11um - 3.7um Product 1km (C)/satellite-2014-04-21-04/satellite-2014-04-21-
04.bin.1
satellite/MODIS 11um - 3.7um Product 1km (C)/satellite-2014-04-21-04/satellite-2014-04-21-
04.h5
satellite/East CONUS/Imager Visible/satellite-2014-04-21-03/satellite-2014-04-21-03.bin.1
satellite/East CONUS/Imager Visible/satellite-2014-04-21-03/satellite-2014-04-21-03.h5
```

Additional Example 6. Raw Data: manual/obs.

One WFO had some data from the manual endpoint processed by their obs plugin. This data was stored in /data_store/manual/obs:

```
manual/obs/20140421/00/awomtn.142341234
manual/obs/20140421/01/awomtn.142341234
manual/obs/20140421/01/awowtn.142341234
manual/obs/20140421/03/awowtn.142341234
manual/obs/20140421/04/awomtn.142341234
manual/obs/20140421/04/awowtn.142341234
manual/obs/20140421/06/awowtn.142341234
manual/obs/20140421/07/awomtn.142341234
manual/obs/20140421/07/awowtn.142341234
manual/obs/20140421/08/awowtn.142341234
manual/obs/20140421/09/awomtn.142341234
manual/obs/20140421/09/awowtn.142341234
manual/obs/20140421/18/awowtn.142341234
```

The filenames do not contain time, other than the time in the directory that the manual endpoint supplies. The two options are to use the directory time or the file write time. For performance reasons, this example uses the directory time.

No baseline RAW_DATA.xml entry handles obs data from /data_store/manual.

```
<dataSet>
   <dirPattern>(manual)/(obs)/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
```

```
        <displayLabel>{1}-{2}</displayLabel>
        <dateGroupIndices>3,4,5,6</dateGroupIndices>
    </dataSet>
```

Sample File:

```
manual/obs/20140421/08/awowtn.142341234
```

Additional Example 7. Raw Data: SHEF from WFO HUN.

While these filenames are different, the date/time specifications (or lack thereof) and the resulting dataset entry is the same as shown in Data Example 10, so this is a similar entry to that used in that example.

```
manual/shef/20140421/00/HCNShef.dat
manual/shef/20140421/00/uahcoop6hrly.dat
manual/shef/20140421/06/uahcoop6hrly.dat
manual/shef/20140421/06/uahcoop24hrly.dat.1398060993
manual/shef/20140421/12/HCNShef.dat
manual/shef/20140421/12/uahcoop6hrly.dat
manual/shef/20140421/14/HUNCOCOSH.1398091463
manual/shef/20140421/18/HUNCOCOSH.1398103163
manual/shef/20140421/18/uahcoop6hrly.dat
manual/shef/20140420/23/HUNCOCOSH.1398037481
```

```
    <dataSet>
        <dirPattern>(manual)/(shef)/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
        <displayLabel>{1}-{2}</displayLabel>
        <dateGroupIndices>3,4,5,6</dateGroupIndices>
    </dataSet>
```

Additional Example 8. Raw Data: unknown

The Milwaukee WFO has some SHEF data that originates from CoCoRAHS. The algorithm the manual endpoint uses to reorganize the data (as it moves the data from /awips2/edex/data/manual to subdirectories /data_store/manual) did not know how to handle this data, so the manual endpoint filed it in an unknown directory.

```
manual/unknown/20140421/12/CoCoShef.snow.IN.txt.1398081727
manual/unknown/20140421/13/CoCoShef.snow.IL.txt.1398085266
manual/unknown/20140421/13/CoCoShef.pcpn.WI.txt.1398085286
manual/unknown/20140421/13/CoCoShef.pcpn.IN.txt.1398085306
manual/unknown/20140421/13/CoCoShef.snow.IN.txt.1398085326
manual/unknown/20140421/13/CoCoShef.pcpn.IL.txt.1398085326
manual/unknown/20140421/13/CoCoShef.snow.WI.txt.1398085346
manual/unknown/20140421/15/CoCoShef.pcpn.WI.txt.1398094279
manual/unknown/20140421/15/CoCoShef.snow.IL.txt.1398094279
manual/unknown/20140421/15/CoCoShef.pcpn.IL.txt.1398094319
manual/unknown/20140421/15/CoCoShef.snow.IN.txt.1398094319
manual/unknown/20140421/15/CoCoShef.pcpn.IN.txt.1398094319
manual/unknown/20140421/15/CoCoShef.snow.WI.txt.1398094339
```

```
manual/unknown/20140421/17/CoCoShef.snow.IL.txt.1398099643
manual/unknown/20140421/17/CoCoShef.pcpn.WI.txt.1398099663
manual/unknown/20140421/17/CoCoShef.pcpn.IN.txt.1398099683
manual/unknown/20140421/17/CoCoShef.pcpn.IL.txt.1398099703
manual/unknown/20140421/17/CoCoShef.snow.WI.txt.1398099703
manual/unknown/20140421/17/CoCoShef.snow.IN.txt.1398099703
```

The Epoch seconds in the file extension could be used for this data, but that time is roughly equivalent to the directory time, so for performance reasons, the entry below uses the directory time.

```
<dataSet>
   <dirPattern>(manual)/(unknown)/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <displayLabel>{1}-{2}</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample File:

```
manual/unknown/20140421/15/CoCoShef.pcpn.IL.txt.1398094319
```

Alternatively, the <displayLabel> could be modified to explicitly say this is CoCoRAHS data, but this entry is generic enough that it could potentially archive data other than CoCoRAHS that could also be filed in /data_store/manual/unknown by the manual endpoint.

Additional Example 9. Raw Data: /data_store/manual/grib from WFO HUN (GOES-R CI and NASA SPoRT LIS)

Two types of files from the University of Alabama-Huntsville (UAH) and NASA SPoRT stand out from a two-hour sample of data from /data_store/manual/grib. These both have valid date/times in YYYYMMDD_HHMM format. These samples also indicate that the valid times can differ from the directory times. The <filePattern> tag used below discriminates between the UAH Convective Initiation nowcast and the sportlis (SPoRT Land Information System), allowing them to be archived separately, if desired, by the user. Separate entries are required because <filePattern> expressions can't be used as part of the <displayLabel>.

```
manual/grib/20140420/22/20140420_2200_UAH_GOESRCI_wconus_Nowcst.grb2
manual/grib/20140420/22/20140420_2202_UAH_GOESRCI_econus_Nowcst.grb2
manual/grib/20140420/22/20140420_2211_UAH_GOESRCI_wconus_Nowcst.grb2
manual/grib/20140420/22/20140420_2215_UAH_GOESRCI_wconus_Nowcst.grb2
manual/grib/20140420/22/20140420_2215_UAH_GOESRCI_econus_Nowcst.grb2
manual/grib/20140420/22/sportlis_seus_awips_20140419_0000.grb2.1398033237
manual/grib/20140420/22/sportlis_seus_awips_20140419_0600.grb2.1398033297
manual/grib/20140420/22/sportlis_seus_awips_20140420_0000.grb2.1398033477
manual/grib/20140420/22/20140420_2230_UAH_GOESRCI_wconus_Nowcst.grb2
manual/grib/20140420/22/sportlis_seus_awips_20140420_0600.grb2.1398033557
manual/grib/20140420/22/sportlis_seus_awips_20140420_1200.grb2.1398033617
manual/grib/20140420/22/sportlis_seus_awips_20140420_1800.grb2.1398033677
manual/grib/20140420/22/20140420_2232_UAH_GOESRCI_econus_Nowcst.grb2
manual/grib/20140420/22/sportlis_seus_awips_20140421_0000.grb2.1398033757
manual/grib/20140420/22/sportlis_seus_awips_20140421_1800.grb2.1398033938
manual/grib/20140420/22/20140420_2241_UAH_GOESRCI_wconus_Nowcst.grb2
```

```
manual/grib/20140420/22/20140420_2245_UAH_GOESRCI_wconus_Nowcst.grb2
manual/grib/20140420/23/20140420_2245_UAH_GOESRCI_econus_Nowcst.grb2
manual/grib/20140420/23/20140420_2300_UAH_GOESRCI_wconus_Nowcst.grb2
manual/grib/20140420/23/20140420_2302_UAH_GOESRCI_econus_Nowcst.grb2
manual/grib/20140420/23/20140420_2311_UAH_GOESRCI_wconus_Nowcst.grb2
manual/grib/20140420/23/20140420_2315_UAH_GOESRCI_wconus_Nowcst.grb2
manual/grib/20140420/23/20140420_2315_UAH_GOESRCI_econus_Nowcst.grb2
manual/grib/20140420/23/20140420_2330_UAH_GOESRCI_wconus_Nowcst.grb2
```

UAH Convective Initiation Product:

```
<dataSet>
   <dirPattern>(manual)/grib/\d{8}/\d{2}</dirPattern>
   <filePattern>.*(\d{4})(\d{2})(\d{2})_(\d{2})(\d{2}).*(GOESRCI).*</filePattern>
   <displayLabel>{1}-GOES-R CI</displayLabel>
   <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

Sample File:

```
manual/grib/20140420/22/20140420_2241_UAH_GOESRCI_wconus_Nowcst.grb2
```

NASA SPoRT Land Information System grids:

```
<dataSet>
   <dirPattern>(manual)/grib/\d{8}/\d{2}</dirPattern>
   <filePattern>.*(sportlis).*(\d{4})(\d{2})(\d{2})_(\d{2})(\d{2}).*</filePattern>
   <displayLabel>{1}-SPoRT Land Info Sys</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample File:

```
manual/grib/20140420/22/sportlis_seus_awips_20140421_1800.grb2.1398033938
```

Additional Example 10. Raw Data: Ensembles from narre

At least one WFO in Central Region processed grib files from the narre system.

```
grib/20140421/19/LDAD-GRIB-cr.narre.t18z.mean.grd130.f02.grib2_2014-04-21_194008
grib/20140421/19/LDAD-GRIB-cr.narre.t18z.mean.grd130.f01.grib2_2014-04-21_194008
grib/20140421/19/LDAD-GRIB-cr.narre.t18z.mean.grd130.f03.grib2_2014-04-21_194008
grib/20140421/19/LDAD-GRIB-cr.narre.t18z.spread.grd130.f01.grib2_2014-04-21_194033
grib/20140421/19/LDAD-GRIB-cr.narre.t18z.mean.grd130.f07.grib2_2014-04-21_194036
grib/20140421/19/LDAD-GRIB-cr.narre.t18z.prob.grd130.f01.grib2_2014-04-21_194038
grib/20140421/19/LDAD-GRIB-cr.narre.t18z.spread.grd130.f05.grib2_2014-04-21_194058
grib/20140421/19/LDAD-GRIB-cr.narre.t18z.spread.grd130.f07.grib2_2014-04-21_194058
```

```
grib/20140421/20/LDAD-GRIB-cr.narre.t19z.mean.grd130.f06.grib2_2014-04-21_204057
grib/20140421/20/LDAD-GRIB-cr.narre.t19z.mean.grd130.f08.grib2_2014-04-21_204057
grib/20140421/20/LDAD-GRIB-cr.narre.t19z.prob.grd130.f02.grib2_2014-04-21_204057
grib/20140421/20/LDAD-GRIB-cr.narre.t19z.prob.grd130.f06.grib2_2014-04-21_204057
grib/20140421/20/LDAD-GRIB-cr.narre.t19z.prob.grd130.f04.grib2_2014-04-21_204057
grib/20140422/00/LDAD-GRIB-cr.narre.t23z.mean.grd130.f01.grib2_2014-04-22_004033
```

From this two hour sample of data, it is apparent that the model time ("t##z") only contains the hour of the day. There is an hour lag between the directory time (or LDAD time at the end of the file) and the model time. In this case, we can't use two digit hour from the filename because the 23Z file (at the day changeover) would be incorrectly dated and not pulled correctly (as can be seen by the last filename above; if we used the directory date and the 23z from the filename, the archiver would think the file represented 23Z on April 22 rather than 23Z on the 21st). Here's a situation where the archived data will always be organized by the directory time and will be an hour off of the model run time. Little can be done about this situation unless the filenames coming through LDAD could be constructed differently.

```
<dataSet>
   <dirPattern>(manual)/grib/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <filePattern>.*(narre).*</filePattern>
   <displayLabel>{1}-narre</displayLabel>
   <dateGroupIndices>2,3,4,5</dateGroupIndices>
 </dataSet>
```

Sample File:

```
manual/grib/20140422/00/LDAD-GRIB-cr.narre.t23z.mean.grd130.f01.grib2_2014-04-22_004033
```

Additional Example 11. Raw Data: WRF Model from Alaska Region

WFOs in the Alaska Region receive a specific version of the WRF model. In this sample of data, the time of the model run is given as the Julian date, followed by the forecast hour. "201411321001700" equates to the 113th day of 2014 (April 23rd) at 2100 UTC for a forecast time of 17:00.

```
grib/20140423/22/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411321001300.grib2.Grib2F.1398293745

grib/20140423/22/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411321001500.grib2.Grib2F.1398293765

grib/20140423/22/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411321001400.grib2.Grib2F.1398293765

grib/20140423/22/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411321001600.grib2.Grib2F.1398293785

grib/20140423/22/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411321001800.grib2.Grib2F.1398293805
```

```
grib/20140423/22/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411321001700.grib2.Grib2F.1398293805

grib/20140424/02/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411401000000.grib2.Grib2F.1398307861

grib/20140424/02/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411401000100.grib2.Grib2F.1398307881

grib/20140424/02/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411401000200.grib2.Grib2F.1398307901

grib/20140424/02/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411401000300.grib2.Grib2F.1398307921

grib/20140424/02/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411401000400.grib2.Grib2F.1398307941
```

The archiver configuration files can handle a <timeType> value of Julian with three <dateGroupIndices> (year, day, hour) rather than four (year, month, day, hour).

```
    <dataSet>
        <dirPattern>(manual)/grib/\d{8}/\d{2}/.*/.*</dirPattern>
        <filePattern>.*(WRFRR).*(\d{4})(\d{3})(\d{2})(\d{6}).*</filePattern>
        <displayLabel>{1}-WRFRR</displayLabel>
        <timeType>Julian</timeType>
        <dateGroupIndices>3,4,5</dateGroupIndices>
    </dataSet>
```

Sample File:

```
manual/grib/20140424/02/grib/grib2LargeSplit/LDADGRIB-
WRFRR_AFG_201411401000300.grib2.Grib2F.1398307921
```

Additonal Example 12. SPC Lightning for Alaska Region

Alaska region WFOs receive a special lightning product from the Storm Prediction Center. Here is a listing of a couple data files. The data time is represented by the YYYYMMDD_HHMM in the filename.

```
grib/20140423/19/grib/grib2LargeSplit/LDADGRIB-
SPCltng_vrh_ltggfsgrb_20140423_1200.grib2.Grib2F.1398280844

grib/20140424/01/grib/grib2LargeSplit/LDADGRIB-
SPCltng_vrh_ltggfsgrb_20140423_1800.grib2.Grib2F.1398302441
```

```
    <dataSet>
      <dirPattern>(manual)/grib/\d{8}/\d{2}/.*/.*</dirPattern>
      <filePattern>.*(SPCltng_vrh).*(\d{4})(\d{2})(\d{2})_(\d{2}).*</filePattern>
      <displayLabel>{1}-SPC Lightning (Alaska Region)</displayLabel>
      <dateGroupIndices>3,4,5,6</dateGroupIndices>
    </dataSet>
```

Sample File:

```
manual/grib/20140424/13/grib/grib2LargeSplit/LDADGRIB-
SPCltng_vrh_ltggfsgrb_20140424_0600.grib2.Grib2F.1398345643
```

Additional Example 13. HPC QPF NDFD grids

```
manual/grib/20131119/00/hpcqpf67.hpcqpf67_p06m_2013111900f126.Grb2
manual/grib/20131119/00/hpcqpf67.hpcqpf67_p06m_2013111900f132.Grb2
manual/grib/20131119/00/hpcqpf67.hpcqpf67_p06m_2013111900f138.Grb2
manual/grib/20131119/00/hpcqpf67.hpcqpf67_p06m_2013111900f144.Grb2
manual/grib/20131119/00/hpcqpf67.hpcqpf67_p06m_2013111900f150.Grb2
manual/grib/20131119/00/hpcqpf67.hpcqpf67_p06m_2013111900f156.Grb2
```

The filename structures are **manual/grib/YYYYMMDD/HH/hpcqpf67.hpcqpf67_p06m.YYYYMMDDfhhh.Grb2.** The full date is included in both the directory and the filenames, and both agree, so the time specification can come from the directory.

```
<dataSet>
   <dirPattern>(manual)/grib/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <filePattern>(hpcqpf67).*</filePattern>
   <timeType>Date</timeType>
   <displayLabel>{1}-hpcqpf67</displayLabel>
   <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

Sample File:

```
manual/grib/20131119/00/hpcqpf67.hpcqpf67_p06m_2013111900f126.Grb2
```

Additional Example 14. HRRR Model

```
manual/grib/20140303/21/LDAD-
GRIB.HRRR.003Hour.DZDT.50_to_80_Sigma.201403032000.grb2_2014-03-03_213432
manual/grib/20140303/21/LDAD-GRIB.HRRR.002Hour.HGT.ADCL.201403032000.grb2_2014-
03-03_213432
manual/grib/20140303/21/LDAD-GRIB.HRRR.003Hour.RVIL.EA.201403032000.grb2_2014-03-
03_213433
manual/grib/20140303/21/LDAD-GRIB.HRRR.003Hour.HGT.ADCL.201403032000.grb2_2014-
03-03_213453
manual/grib/20140303/21/LDAD-
GRIB.HRRR.004Hour.CAPE.Surface.201403032000.grb2_2014-03-03_213713
manual/grib/20140303/21/LDAD-
GRIB.HRRR.004Hour.DPT.2m_FHAG.201403032000.grb2_2014-03-03_213713
manual/grib/20140303/21/LDAD-
GRIB.HRRR.004Hour.TMP.Surface.201403032000.grb2_2014-03-03_213713
```

```
manual/grib/20140303/21/LDAD-GRIB.HRRR.004Hour.CNVIRF.EA.201403032000.grb2_2014-03-03_213713
manual/grib/20140303/21/LDAD-GRIB.HRRR.004Hour.MDNDFT.40000Pa_to_100000Pa_Isobaric.201403032000.grb2_2014-03-03_213713
```

The HRRR data have this filename structure:
manual/grib/YYYYMMDD/HH/LDAD-GRIB.HRRR.FFFHour.{ModelField}.{ModelLayer}.YYYYMMDDHHMM.grb2_YYYY-MM-DD_HHMMSS, where the date/time in the directory is the time the AWIPS-2 manual endpoint received the data, FFF is the forecast hour, the YYYYMMDDHHMM is the model run time, and the _YYYY-MM-DD_HHMMSS is the time LDAD received the data. Thus, the first and last times (both times of receipt) should be similar. The HRRR data need to use the date of the model run, which is the first date in the filename.

```
<dataSet>
   <dirPattern>(manual)/grib/\d{8}/\d{2}</dirPattern>
   <filePattern>.*(HRRR).*(\d{4})(\d{2})(\d{2})(\d{2}).*</filePattern>
   <timeType>Date</timeType>
   <displayLabel>{1}-HRRR</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample File:

```
manual/grib/20140303/16/LDAD-GRIB.HRRR.002Hour.TMP.Surface.201403031500.grb2_2014-03-03_163312
```

Additional Example 15. NAM80 Model (beyond 60 hours plus 6Z and 18Z runs)

Several WFOs receive data files that look similar to these:

```
manual/grib/20140303/20/LDAD-GRIB.2014030318.GribF66_2014-03-03_204545
manual/grib/20140303/20/LDAD-GRIB.2014030318.GribF84_2014-03-03_204605
manual/grib/20140304/02/LDAD-GRIB.2014030400.GribF66_2014-03-04_024548
manual/grib/20140304/02/LDAD-GRIB.2014030400.GribF78_2014-03-04_024549
manual/grib/20140304/02/LDAD-GRIB.2014030400.GribF72_2014-03-04_024549
manual/grib/20140304/02/LDAD-GRIB.2014030400.GribF84_2014-03-04_024549
manual/grib/20140304/08/LDAD-GRIB.2014030406.GribF72_2014-03-04_084600
```

Further examination of the data revealed these files contain additional output grids from the NAM80 model beyond those available in the baseline SBN feed, though the filenames themselves do not describe well this type of data. The filenames are represented by **manual/grib/YYYYMMDD/HH/LDAD-GRIB.YYYYMMDDHH.GribFff_YYYY-MM-DD_HHMMSS**, where ff is the forecast hour.

The archiver needs to use the date of the model run, which is the first date in the filename. ".GribF" followed by two digits apparently is a distinguishing characteristic of these files among other files that appear in /data_store/manual/grib.

```
<dataSet>
   <dirPattern>(manual)/grib/\d{8}/\d{2}</dirPattern>
   <filePattern>.*(\d{4})(\d{2})(\d{2})(\d{2}).GribF(\d{2})_.*</filePattern>
   <timeType>Date</timeType>
   <displayLabel>{1}-NAM80 (Additional Grids)</displayLabel>
   <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

Sample File:

```
manual/ grib/20140303/20/LDAD-GRIB.2014030318.GribF84_2014-03-03_204605
```

Additional Example 16. University of Wisconsin Cloud-Top Cooling/Convective Initiation Products

A number of WFOs receive Convective Initiation products from the Space Science and Engineering Center at the University of Wisconsin.

```
manual/grib/20140303/16/LDAD-
GRIB.SSEC_AWIPS_CONVECT_GRID197_5KM_ALL_GOES_20140303_1645.GRIB2_2014-03-
03_165427
manual/grib/20140303/17/LDAD-
GRIB.SSEC_AWIPS_CONVECT_GRID197_5KM_ALL_GOES_20140303_1715.GRIB2_2014-03-
03_172427
manual/grib/20140303/17/LDAD-
GRIB.SSEC_AWIPS_CONVECT_GRID197_5KM_ALL_GOES_20140303_1732.GRIB2_2014-03-
03_173852
manual/grib/20140303/17/LDAD-
GRIB.SSEC_AWIPS_CONVECT_GRID197_5KM_ALL_GOES_20140303_1745.GRIB2_2014-03-
03_175601
```

The satellite convective initiation products have three date/time specifications, and all three agree, apparently because these are diagnostic rather than prognostic products. Hence, the best date/time to use for this data would be the directory for performance reasons. "SSEC_AWIPS_CONVECT" is a distinguishing characteristic of this data.

```
<dataSet>
   <dirPattern>(manual)/grib/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <filePattern>.*(SSEC_AWIPS_CONVECT).*</filePattern>
   <timeType>Date</timeType>
   <displayLabel>{1}-SSEC Convective Initiation</displayLabel>
   <dateGroupIndices>2,3,4,5</dateGroupIndices>
</dataSet>
```

Sample file: manual /grib/20140304/14/LDAD-
GRIB.SSEC_AWIPS_CONVECT_GRID197_5KM_ALL_GOES_20140304_1432.GRIB2_2014-03-04_143850

Additional Example 17. Raw Data: Local WRF model at WFO OUN

The Norman WFO utilizes a best practice in that they store their local data in /data_store/local, rather than /data_store/manual. They use a Python script to notify EDEX to ingest the data by placing a notification message on QPID's external.dropbox queue. In this manner, they are able to better organize their local data. There are no disagreements between their directory times and their data valid times. Hence we can use the directory times and gain a performance boost. For more information on OUN's script, see the "LDAD Ingest VIA QPID" page on the AWIPS II Site Configuration Wiki at https://collaborate.nws.noaa.gov/trac/siteconfig/wiki/LdadIngestQpid.

```
local/ounwrf/20140420/2000/wrfoun41404202000_arw.Grb2F0000.1398028916
local/ounwrf/20140420/2000/wrfoun41404202000_arw.Grb2F0015.1398028996
local/ounwrf/20140420/2000/wrfoun41404202000_arw.Grb2F0030.1398029056
local/ounwrf/20140420/2000/wrfoun41404202000_arw.Grb2F0045.1398029136
local/ounwrf/20140420/2000/wrfoun41404202000_arw.Grb2F0745.1398031327
local/ounwrf/20140420/2000/wrfoun41404202000_arw.Grb2F0800.1398031398
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0000.1398036322
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0015.1398036402
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0030.1398036462
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0045.1398036542
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0100.1398036602
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0115.1398036683
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0130.1398036743
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0145.1398036823
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0200.1398036883
```

```
<dataSet>
   <dirPattern>(local)/(ounwrf)/(\d{4})(\d{2})(\d{2})/(\d{2})(\d{2})</dirPattern>
   <displayLabel>{1}-{2}</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample File:

```
local/ounwrf/20140420/2200/wrfoun41404202200_arw.Grb2F0030.1398036462
```

Additional Example 18. Raw Data: MRMS data at WFO OUN

The Norman WFO receives gridded "Multi-Radar Multi-Sensor" products from the National Severe Storms Laboratory (NSSL). Like the OUN WRF, they also organize these in directories that agree with the valid time.

```
local/mrms/20140421/18/MESH_Max_120min_00.00_20140421-185837.grib2.1398106734
local/mrms/20140421/18/Reflectivity_0C_00.00_20140421-185837.grib2.1398106734
local/mrms/20140421/18/H50_Above_H253_00.00_20140421-185837.grib2.1398106734
```

```
local/mrms/20140421/18/Reflectivity_0C_00.00_20140421-185932.grib2.1398106774
local/mrms/20140421/18/EchoTop_50_00.00_20140421-185932.grib2.1398106774
local/mrms/20140421/18/Reflectivity_-20C_00.00_20140421-185932.grib2.1398106774
local/mrms/20140421/18/H50_Above_H253_00.00_20140421-185932.grib2.1398106774
local/mrms/20140421/18/MESH_00.00_20140421-185932.grib2.1398106774
local/mrms/20140421/19/MESH_Max_60min_00.00_20140421-190036.grib2.1398106854
local/mrms/20140421/19/MESH_00.00_20140421-190036.grib2.1398106854
local/mrms/20140421/19/H50_Above_H253_00.00_20140421-190036.grib2.1398106854
local/mrms/20140421/19/Reflectivity_0C_00.00_20140421-190036.grib2.1398106854
local/mrms/20140421/19/MESH_Max_120min_00.00_20140421-190036.grib2.1398106854
local/mrms/20140421/19/Reflectivity_-20C_00.00_20140421-190036.grib2.1398106854
local/mrms/20140421/19/EchoTop_50_00.00_20140421-190036.grib2.1398106854
local/mrms/20140421/19/H50_Above_H253_00.00_20140421-190126.grib2.1398106894
local/mrms/20140421/19/EchoTop_50_00.00_20140421-190126.grib2.1398106894
local/mrms/20140421/19/Reflectivity_-20C_00.00_20140421-190126.grib2.1398106894
local/mrms/20140421/19/MESH_00.00_20140421-190126.grib2.1398106894
```

```
<dataSet>
   <dirPattern>(local)/(mrms)/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <displayLabel>{1}-{2}</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample File:

```
local/mrms/20140421/19/Reflectivity_-20C_00.00_20140421-190126.grib2.1398106894
```

As you can see, many MRMS grids are produced every few minutes. (This listing only includes a subset of available MRMS grids.) In addition, the number of MRMS products is likely to increase. Therefore, the user may wish to archive certain types of MRMS products based on the type of event. Below are two examples, one that specifies only the Reflectivity grids and another example that will archive only the MESH (Maximum Expected Hail Size) grids. Note that the .* prior to (Reflectivity) and (MESH) do not match anything (.* can match 0 or more characters), but this .* is included in case other information may be prepended to the filename in the future.

Reflectivity Grids:

```
<dataSet>
   <dirPattern>(local)/(mrms)/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <filePattern>.*(Reflectivity).*</filePattern>
   <displayLabel>{1}-{2} Reflectivity</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample File:

```
local/mrms/20140421/19/Reflectivity_-20C_00.00_20140421-190126.grib2.1398106894
```

Maximum Expected Hail Size (MESH) Grids:

```
<dataSet>
  <dirPattern>(local)/(mrms)/(\d{4})(\d{2})(\d{2})/(\d{2})</dirPattern>
   <filePattern>.*(MESH).*</filePattern>
  <displayLabel>{1}-{2} MESH</displayLabel>
   <dateGroupIndices>3,4,5,6</dateGroupIndices>
</dataSet>
```

Sample File:

```
local/mrms/20140421/19/MESH_00.00_20140421-190036.grib2.1398106854
```